Why Does the Kronecker Model Result in Misleading Capacity Estimates?

Vasanthan Raghavan*, Jayesh H. Kotecha, and Akbar M. Sayeed

Abstract

Many recent works that study the performance of multi-input multi-output (MIMO) systems in practice assume a Kronecker model where the variances of the channel entries, upon decomposition on to the transmit and the receive eigen-bases, admit a separable form. Measurement campaigns, however, show that the Kronecker model results in poor estimates for capacity. Motivated by these observations, a channel model that does not impose a separable structure has been recently proposed and shown to fit the capacity of measured channels better. In this work, we show that this recently proposed modeling framework can be viewed as a natural consequence of channel decomposition on to its *canonical coordinates*, the transmit and/or the receive eigen-bases. Using tools from random matrix theory, we then establish the theoretical basis behind the Kronecker mismatch at the low- and the high-SNR extremes: 1) *Sparsity* of the dominant statistical degrees of freedom (DoF) in the true channel at the low-SNR extreme, and 2) *Non-regularity* of the sparsity structure (disparities in the distribution of the DoF across the rows and the columns) at the high-SNR extreme.

Index Terms

Correlation, fading channels, information rates, MIMO systems, multiplexing, random matrix theory, sparse systems.

I. INTRODUCTION

Under the assumption of spatially independent and identically distributed (i.i.d.) Rayleigh fading between antenna pairs, multi-input multi-output (MIMO) systems achieve a linear growth

in multiplexing gain and coherent capacity with the number of antennas [1], [2]. However, the rich scattering assumption is idealistic and most physical channels encountered in practice exhibit clustered scattering and spatially correlated links [3]–[5]. Correlated MIMO channels have been theoretically studied mainly in the contexts of the separable correlation model (also known as the *Kronecker model*) [3], [4], and the virtual representation framework for uniform linear arrays (ULAs) [6]–[9]. The Kronecker model assumes separability in correlation induced by the transmitter and the receiver arrays which limits the degrees of freedom (DoF) in modeling the channel. Though this model has been shown to be accurate in certain settings (especially 2×2 scenarios) [10]–[14], the separability assumption limits its applicability to more realistic settings where the gains accrued with MIMO make it a viable choice. The virtual representation does not assume such separability, but is applicable only for ULAs.

Contributions:

- i) In this paper, we develop a unified statistical modeling framework for Rayleigh fading MIMO channels based on decomposition of the channel matrix on its *canonical coordinates*, the transmit and/or the receive eigen-bases. Motivated by virtual representation, these models do not assume separable statistics and are applicable to general array geometries. Like the Kronecker model, the eigen-modes of the scattering environment decide the transmit and the receive eigen-bases whereas the canonical channel matrix embodies the statistically independent DoF that govern channel capacity and diversity.

  Depending on the covariance structure of the channel, three models arise in which all the columns/rows/channel entries are uncorrelated. The last case[1], denoted here as the *canonical model* (or CM3 for short), has been proposed as the *Weichselberger model* in [18] and studied from a capacity analysis viewpoint in [19]. The new contribution in this work is the unified development of CM3 as a natural consequence of two other models, denoted as CM1 and CM2. The development of CM1 (and CM2) critically depends on two assumptions about the covariance and the cross-covariance information of the rows (and the columns) of the channel matrix. We establish the criticality of these four assumptions in the development of CM3. To the best of our knowledge, CM1 and CM2 have not been proposed elsewhere in the literature, and could provide useful intermediate models for certain asymmetric MIMO systems.

- ii) Many recent works [18], [20]–[26] have shown that the Kronecker model consistently estimates the capacity of a large class of measured channels poorly and hence they establish the need[2] for more accurate channel modeling. For example, [18], [20]–[22] show that the Kronecker model severely underestimates true capacity whereas under certain conditions,

---

[1]Some of the earlier works of the authors [15], [16] have also suggested this model and developed it independently [17].

[2]In fact, the development of CM3 in [18] is motivated by these observations.

it could also overestimate the true capacity [24], [26]. Nevertheless, motivated by extensive measurement studies, the popular belief is that the "probability of overestimation decreases with increasing antenna number" [26, Footnote 5]. The main focus of this work is to theoretically explain these observations.

Towards this goal, we first note that recent measurement campaigns have also observed that only a few of the statistical DoF are dominant enough to contribute towards reliable communications. That is, measured multi-antenna channels are *sparse* in the canonical domain. Furthermore, the distribution of the sparse DoF across the spatial domain does not observe any *regularity*[3] structure. For example, see [26, Figs. 9 and 11], [20], [21], [28] etc. which plot the sparse, non-regular structure of the *Weichselberger coupling matrix* that reflects the statistical DoF in CM3.
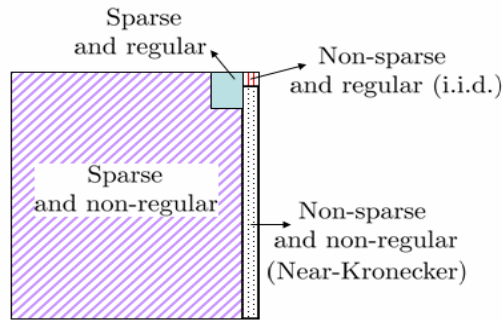

Fig. 1.   Partitioning of the space of all possible channels based on sparsity and regularity.

The foundation for the above experimental evidence lies in theoretical electromagnetic studies that explain sparsity of DoF in different contexts in wireless communications [9], [29]–[35]. Nevertheless, a simpler communication-theoretic motivation for sparsity is that while there may be many channel coefficients whose energy levels are non-zero, they may not be strong enough to be estimated accurately at the transmitter, even statistically. It becomes impossible or too costly to estimate such coefficients accurately and thus from the transmitter's viewpoint, it is reasonable to treat their contributions as noise. Thus, we can partition the space of all possible channels into four classes, as in Fig. 1, with the class of sparse and non-regular channels being the most predominant. In this work, we develop a mathematical framework for probabilistically modeling sparse multi-antenna

---

[3]Let $\mathbf{H}_c$ be an $N_r \times N_t$ random matrix with independent entries and let the variance of $\mathbf{H}_c[i,j]$ be given by $\mathbf{P}_c[i,j]$. A channel is called *column-regular* if $\sum_{i=1}^{N_r} \mathbf{P}_c[i,j]$ is equal for all j, *row-regular* if the above condition is true for $\mathbf{H}_c^T$, and *regular* if it is both row- and column-regular [27]. Otherwise, it is *non-regular*.

channels. The framework developed here allows us to adjust the average number of dominant channel coefficients and theoretically study the impact of a Kronecker model on the capacity mismatch.

- iii) We assume that the statistics of the true channel in the canonical domain has a sparse, non-separable structure. Based on recent works (see [36] and references therein) that establish the accuracy of a Gaussian approximation to outage capacity, we also assume that the ergodic capacity ($C_{\mathsf{erg}}(\rho)$) and the variance of capacity ($V(\rho)$) are the key figures-of-merit. The main results of this work are obtained for the low- and the high-SNR extremes in the large-system (antenna) regime, and are summarized in Table I.

Table I: Summary of Main Results		
	Mismatch Metric	Conclusions
$\rho \to 0$	$C_{\mathsf{erg, can}}(\rho) = C_{\mathsf{erg, kron}}(\rho)$ for all chan. $\frac{V_{\mathsf{can}}(\rho)}{V_{\mathsf{kron}}(\rho)} \to \frac{1}{p}\left(1 + \frac{\sigma^2}{\mu^2}\right)$ where $\{p, \mu, \sigma\}$ are sparse model parameters (See Theorem 2 and Prop. 2 for details)	I.I.D. $\Longleftarrow \mathbf{H}_c \Longrightarrow$ Sparse Decreases $\Longleftarrow$ Mism. $\Longrightarrow$ Increases
$\rho \to \infty$	$C_{\mathsf{erg, can}}(\rho) - C_{\mathsf{erg, kron}}(\rho) \to$ $\frac{N}{2}\log_2\left(\frac{\mathrm{AM}_{\mathsf{row\ pow}} \cdot \mathrm{AM}_{\mathsf{col\ pow}}}{\mathrm{GM}_{\mathsf{row\ pow}} \cdot \mathrm{GM}_{\mathsf{col\ pow}}}\right)$ where $\mathrm{AM}_\bullet$ and $\mathrm{GM}_\bullet$ are arithmetic and geometric means of row and column powers of the true channel (See Theorem 4 for details)	Regular $\Longleftarrow \mathbf{H}_c \Longrightarrow$ Non-regular Decreases $\Longleftarrow$ Mism. $\Longrightarrow$ Increases

We show that for *almost every* sparse channel: a) Using marginal sum statistics to generate a Kronecker fit results in an artificial increase in the number of DoF, b) As a result, the channel power is spread across the increased DoF, and c) Hence, the Kronecker model offers a poor estimate for capacity. Towards establishing this connection, we develop a tight approximation for the mean of the log-determinant of random matrices (with independent entries) which is of independent interest in MIMO analysis and design.

In the high-SNR extreme, the Kronecker model *underestimates* $C_{\mathsf{erg}}(\rho)$ for all channels. The level of underestimation decreases as the channel becomes more regular. In the low-SNR extreme, $C_{\mathsf{erg}}(\rho)$ is the same with either model. The Kronecker model underestimates $V(\rho)$ with the level of underestimation decreasing as the channel becomes less sparse. Thus, for a large class of channels that are sparse and non-regular, the Kronecker model *underestimates* the outage capacity at all reliability levels (and also the reliability at all

data rates) in the medium- to high-SNR regime. On the other hand, for any channel in the low-SNR regime, and regular channels in the medium- to high-SNR regime, the Kronecker model *overestimates* capacity at high levels of reliability (and reliability at low data rates), and *vice versa*.

Organization: This paper is organized as follows. The canonical statistical modeling framework for correlated multi-antenna channels is developed in Section II with the key properties of the proposed model elucidated in Section III. In Section IV, we explore practical modeling issues and show how the canonical and the Kronecker models are used to describe realistic measured channels. A brief summary of MIMO capacity issues is provided in Section V with a comparative study of the two models performed in Section VI. Conclusions are drawn in Section VII.

Notation: We use upper-case and lower-case bold symbols for matrices and vectors, respectively. If $\mathbf{X}$ is an $M \times N$ matrix, $\mathbf{x} = \text{vec}(\mathbf{X})$ denotes the $MN \times 1$ vector obtained by stacking columns of $\mathbf{X}$. The entry in the m-th row and n-th column, and the m-th diagonal entry of $\mathbf{X}$ are denoted by $\mathbf{X}[m, n]$ and $\mathbf{X}[m] = \mathbf{X}[m, m]$, respectively. The complex conjugate, regular transpose and Hermitian transpose of $\mathbf{X}$ are denoted by $\mathbf{X}^*, \mathbf{X}^T$ and $\mathbf{X}^H$ while its inverse, trace and determinant are denoted by $\mathbf{X}^{-1}$, $\text{Tr}(\mathbf{X})$ and $\det(\mathbf{X})$, respectively. The operators $E[\cdot]$, $\otimes$ and $\odot$ stand for expectation, Kronecker and Hadamard products. The indicator function of a set $\mathcal{A}$ and its probability are given by $\chi(\mathcal{A})$ and $\text{Pr}(\mathcal{A})$. We use the standard big-Oh ($\mathcal{O}$) and little-oh (o) notations, $\sim$ for equality in distribution, and $X \sim \mathcal{CN}(\mu, \sigma^2)$ to indicate that X is a complex Gaussian random variable with mean μ and variance σ^2.

II. CANONICAL MODELING OF CORRELATED MIMO CHANNELS

Consider a narrowband, Rayleigh fading MIMO channel with N_t transmit and N_r receive antennas. The $N_r \times 1$ received vector $\mathbf{y}$ is related to the $N_t \times 1$ transmit vector $\mathbf{x}$ by

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n} \tag{1}$$

where $\mathbf{H}$ is the $N_r \times N_t$ channel matrix and $\mathbf{n}$ is the independent, white Gaussian noise added at the receiver. The entries of $\mathbf{H}$ are zero mean, complex Gaussian that satisfy

$$\mathbf{h} \triangleq \text{vec}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}) \tag{2}$$

for some positive semi-definite channel covariance matrix $\mathbf{R}$.

We now describe three canonical decompositions of MIMO channels. Let $\mathbf{Q}_t \triangleq E\left[\mathbf{H}^H \mathbf{H}\right]$ and $\mathbf{Q}_r \triangleq E\left[\mathbf{H}\mathbf{H}^H\right]$ denote the transmit and the receive covariance matrices. Let their respective eigen-decompositions be given by $\mathbf{Q}_t = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{U}_t^H$ and $\mathbf{Q}_r = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H$ where the columns of $\mathbf{U}_t$ and $\mathbf{U}_r$ are eigenvectors of $\mathbf{Q}_t$ and $\mathbf{Q}_r$ with the corresponding eigenvalues denoted by diagonal entries of $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_r$.

Lemma 1: Any channel matrix $\mathbf{H}$ can be written in the canonical form: $\mathbf{H} = \mathbf{H}_t \mathbf{U}_t^H$ such that $E\left[\mathbf{H}_t^H \mathbf{H}_t\right] = \mathbf{\Lambda}_t$ and

$$\mathbf{h}_t \triangleq \text{vec}(\mathbf{H}_t) = \text{vec}(\mathbf{H}\mathbf{U}_t) = (\mathbf{U}_t^T \otimes \mathbf{I})\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_t) \tag{3}$$

$$\mathbf{R}_t \triangleq E\left[\mathbf{h}_t \mathbf{h}_t^H\right] = (\mathbf{U}_t^T \otimes \mathbf{I}) \, \mathbf{R} \, (\mathbf{U}_t^T \otimes \mathbf{I})^H. \tag{4}$$

Similarly, any channel matrix $\mathbf{H}$ can be written in the canonical form: $\mathbf{H} = \mathbf{U}_r \mathbf{H}_r$ such that $E\left[\mathbf{H}_r \mathbf{H}_r^H\right] = \mathbf{\Lambda}_r$ and

$$\mathbf{h}_r \triangleq \text{vec}(\mathbf{H}_r) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_r) \tag{5}$$

$$\mathbf{R}_r \triangleq E\left[\mathbf{h}_r \mathbf{h}_r^H\right] = (\mathbf{I} \otimes \mathbf{U}_r^H) \, \mathbf{R} \, (\mathbf{I} \otimes \mathbf{U}_r). \tag{6}$$

Furthermore, $\mathbf{H}$ can also be written in the canonical form: $\mathbf{H} = \mathbf{U}_r \mathbf{H}_c \mathbf{U}_t^H$ such that $E\left[\mathbf{H}_c^H \mathbf{H}_c\right] = \mathbf{\Lambda}_t$, $E\left[\mathbf{H}_c \mathbf{H}_c^H\right] = \mathbf{\Lambda}_r$ and

$$\mathbf{h}_c \triangleq \text{vec}(\mathbf{H}_c) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_c) \tag{7}$$

$$\mathbf{R}_c \triangleq E\left[\mathbf{h}_c \mathbf{h}_c^H\right] = (\mathbf{U}_t^T \otimes \mathbf{U}_r^H) \, \mathbf{R} \, (\mathbf{U}_t^T \otimes \mathbf{U}_r^H)^H. \tag{8}$$

Proof: The proof is immediate by using the relation: $\text{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \text{vec}(\mathbf{B})$. ∎

It is possible to obtain interesting, yet realistic statistical models that allow tractable performance analysis if we make the following simplifying assumptions. The following notation is used: The vectors $\mathbf{g}_i$ (and $\mathbf{h}_j$) denote the i-th (and the j-th) column of $\mathbf{H}^H$ (and $\mathbf{H}$), respectively, i.e. $\mathbf{H} = [\mathbf{h}_1 \ldots \mathbf{h}_{N_t}] = [\mathbf{g}_1 \ldots \mathbf{g}_{N_r}]^H$.

A. Canonical Model 1 (CM1)

We denote by CM1 a channel that follows the following two assumptions.

Assumption 1: The covariance matrices of all rows of $\mathbf{H}$ have the columns of $\mathbf{U}_t$ as a set of common eigenvectors. That is, $E\left[\mathbf{g}_i \mathbf{g}_i^H\right] = \mathbf{U}_t \mathbf{\Lambda}_t^{ii} \mathbf{U}_t^H$ for some positive semi-definite diagonal matrix $\mathbf{\Lambda}_t^{ii}$.

Assumption 2: The cross-covariance matrices of the rows of $\mathbf{H}$ also have a set of common eigenvectors, given by columns of $\mathbf{U}_t$, i.e. $E\left[\mathbf{g}_i \mathbf{g}_j^H\right] = \mathbf{U}_t \mathbf{\Lambda}_t^{ij} \mathbf{U}_t^H$ for all $i, j, \ i \neq j$ and some diagonal $\mathbf{\Lambda}_t^{ij}$. In general, $\mathbf{\Lambda}_t^{ij}$ need not be positive semi-definite because $E\left[\mathbf{g}_i \mathbf{g}_j^H\right]$ is not Hermitian.

Then, $\mathbf{H}$ can be written as $\mathbf{H} = \mathbf{H}_t \mathbf{U}_t^H$ with the following properties:

- The covariance matrix of each row of $\mathbf{H}_t$ (denoted by $\{\mathbf{g}_{ti}\}$), given by $E\left[\mathbf{g}_{ti} \mathbf{g}_{ti}^H\right]$, is diagonal. This follows directly from Assumption 1 and the fact that $\mathbf{g}_{ti} = \mathbf{U}_t^H \mathbf{g}_i$.
- The columns of $\mathbf{H}_t$ (denoted by $\{\mathbf{h}_{ti}\}$) are uncorrelated with each other, i.e. $E\left[\mathbf{h}_{ti} \mathbf{h}_{tj}^H\right] = \mathbf{0}$ for all $i, j, \ i \neq j$. However, the columns may have arbitrary covariances. This is because the (m, n)-th entry of the cross-covariance between the i-th and the j-th columns of $\mathbf{H}_t$ is

given by $E\left[\mathbf{g}_{tm}^*[i]\mathbf{g}_{tn}[j]\right]$ which can be seen from Assumptions 1 and 2 to be $\delta(i-j)\mathbf{\Lambda}_t^{nm}[j]$ with $\delta(\cdot)$ denoting the Kronecker delta.

We summarize these conclusions in the form of Lemma 2.

*Lemma 2 (*CM1*):* Under Assumptions 1 and 2, any channel can be written as $\mathbf{H} = \mathbf{H}_t\mathbf{U}_t^H$ where the columns of $\mathbf{H}_t$ are uncorrelated with each other. ∎

B. Canonical Model 2 (CM2)

We denote by CM2 a channel which follows the following two assumptions.

Assumption 3: The covariance matrix of the columns of $\mathbf{H}$, given by $E\left[\mathbf{h}_i\mathbf{h}_i^H\right]$, have a set of common eigenvectors, independent of i. We assume that these eigenvectors are columns of $\mathbf{U}_r$.

Assumption 4: The cross-covariance matrices of columns of $\mathbf{H}$ also have a common set of eigenvectors, given by columns of $\mathbf{U}_r$, i.e. $E\left[\mathbf{h}_i\mathbf{h}_j^H\right] = \mathbf{U}_r\mathbf{\Lambda}_r^{ij}\mathbf{U}_r^H$ for all i,j, $i \neq j$ and some diagonal $\mathbf{\Lambda}_r^{ij}$.

*Lemma 3 (*CM2*):* Under Assumptions 3 and 4, any channel can be written as $\mathbf{H} = \mathbf{U}_r\mathbf{H}_r$ where the rows of $\mathbf{H}_r$ are uncorrelated with each other. ∎

It is clear that CM2 is the dual of CM1. In CM1, the transmitter sees parallel channels in the eigen-domain while in CM2, the receiver sees parallel channels in the eigen-domain.

C. Canonical Model 3 (CM3)

We denote a channel which follows Assumptions 1-4 as CM3. This model has been developed independently in [17]–[19].

*Lemma 4 (*CM3*):* Under Assumptions 1-4, any channel can be written as $\mathbf{H} = \mathbf{U}_r\mathbf{H}_c\mathbf{U}_t^H$ where the entries of $\mathbf{H}_c$ are uncorrelated, but not necessarily identically distributed.

Proof: See Appendix A. ∎

Exploiting the fact that the entries of $\mathbf{H}_c$ are uncorrelated under CM3, the channel covariance matrix can be written as

$$\mathbf{R} = (\mathbf{U}_t^T \otimes \mathbf{U}_r^H)^H \mathbf{R}_c (\mathbf{U}_t^T \otimes \mathbf{U}_r^H) \tag{9}$$

where $\mathbf{R}_c$ is diagonal. First, note that the right-hand side of (9) is an eigen-decomposition of $\mathbf{R}$. The matrices $\mathbf{U}_t$ and $\mathbf{U}_r$, which are a set of eigenvector matrices of $\mathbf{Q}_t = E\left[\mathbf{H}^H\mathbf{H}\right]$ and $\mathbf{Q}_r = E\left[\mathbf{H}\mathbf{H}^H\right]$, can be interpreted as transmit and receive eigen-matrices, respectively. Clearly, CM3 is a special case of CM1 (and CM2) where the covariance matrices of the columns (rows) of $\mathbf{H}_t$ ($\mathbf{H}_r$) have the same eigen-matrix $\mathbf{U}_t$ ($\mathbf{U}_r$). In fact, CM3 is an intersection of CM1 and CM2. Our primary focus in the rest of the paper is on CM3, which we will label as the *canonical model*. We also define the *spatial power* matrix $\mathbf{P}_c$ by the relationship

$$\mathbf{P}_c[i,j] \triangleq E\left[|\mathbf{H}_c[i,j]|^2\right] \tag{10}$$

and note that the diagonal entries of $\mathbf{R}_c$ correspond to $\{\mathbf{P}_c[i,j]\}$. Henceforth, we will use this alternate characterization of $\mathbf{R}_c$. We now identify some of the key properties of CM3.

III. PROPERTIES OF THE CANONICAL MODEL

A. Relation to Other Channel Models

We show how two well-known channel models, the Kronecker model and the virtual representation framework, can be regarded as special cases of the canonical model.

1) Kronecker Model: The Kronecker model has been used in [3], [4] and also in many recent works under the assumption that the transmitter and the receiver are surrounded by local scatterers. This model is verified by measurement campaigns for certain environments in [10]–[14]. It assumes separable statistics at the transmitter and the receiver and is given by

$$\mathbf{H} = \boldsymbol{\Sigma}_r^{1/2} \, \mathbf{H}_{\text{iid}} \, \boldsymbol{\Sigma}_t^{1/2}, \qquad \mathbf{R}_k \triangleq E\left[\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H\right] = \boldsymbol{\Sigma}_t \otimes \boldsymbol{\Sigma}_r \tag{11}$$

where the entries of $\mathbf{H}_{\text{iid}}$ are i.i.d. $\mathcal{CN}(0,1)$, and $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\Sigma}_r$ are the transmit and the receive covariance matrices, respectively.

Recall from the discussion in Sec. II that

$$\mathbf{Q}_t = E\left[\mathbf{H}^H\mathbf{H}\right] = \mathbf{U}_t\boldsymbol{\Lambda}_t\mathbf{U}_t^H, \qquad \mathbf{Q}_r = E\left[\mathbf{H}\mathbf{H}^H\right] = \mathbf{U}_r\boldsymbol{\Lambda}_r\mathbf{U}_r^H, \tag{12}$$

$$\mathbf{R} = E\left[\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H\right], \quad \text{and} \quad \text{Tr}(\mathbf{R}) = \text{Tr}(\mathbf{Q}_t) = \text{Tr}(\mathbf{Q}_r). \tag{13}$$

From (11), we then have the following relations for the Kronecker model:

$$\mathbf{Q}_t = \boldsymbol{\Sigma}_t^{1/2} \, E\left[\mathbf{H}_{\text{iid}}^H \, \boldsymbol{\Sigma}_r \, \mathbf{H}_{\text{iid}}\right] \, \boldsymbol{\Sigma}_t^{1/2} = \boldsymbol{\Sigma}_t \, \text{Tr}(\boldsymbol{\Sigma}_r), \tag{14}$$

$$\mathbf{Q}_r = \boldsymbol{\Sigma}_r^{1/2} \, E\left[\mathbf{H}_{\text{iid}} \, \boldsymbol{\Sigma}_t \, \mathbf{H}_{\text{iid}}^H\right] \, \boldsymbol{\Sigma}_r^{1/2} = \boldsymbol{\Sigma}_r \, \text{Tr}(\boldsymbol{\Sigma}_t). \tag{15}$$

The fact that $\mathbf{Q}_\bullet$ and $\boldsymbol{\Sigma}_\bullet$ are scaled versions of each other implies that eigen-decompositions of $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\Sigma}_r$ are given by $\mathbf{U}_t\boldsymbol{\Lambda}_{t,k}\mathbf{U}_t^H$ and $\mathbf{U}_r\boldsymbol{\Lambda}_{r,k}\mathbf{U}_r^H$ where $\boldsymbol{\Lambda}_{t,k} = \boldsymbol{\Lambda}_t/\text{Tr}(\boldsymbol{\Sigma}_r)$ and $\boldsymbol{\Lambda}_{r,k} = \boldsymbol{\Lambda}_r/\text{Tr}(\boldsymbol{\Sigma}_t)$, respectively. Further,

$$\text{Tr}(\mathbf{Q}_t) = \text{Tr}(\mathbf{Q}_r) = \text{Tr}(\boldsymbol{\Sigma}_t)\text{Tr}(\boldsymbol{\Sigma}_r) = \text{Tr}(\mathbf{R}). \tag{16}$$

Thus, we can write the channel in (11) in the CM3 form as

$$\tau\mathbf{H} = \mathbf{U}_r \, \boldsymbol{\Lambda}_r^{1/2} \, \mathbf{U}_r^H \, \mathbf{H}_{\text{iid}} \, \mathbf{U}_t \, \boldsymbol{\Lambda}_t^{1/2} \, \mathbf{U}_t^H = \mathbf{U}_r \, \mathbf{H}_c \, \mathbf{U}_t^H \tag{17}$$

where $\tau = \left[\text{Tr}(\mathbf{R})\right]^{1/2}$. The entries of $\mathbf{H}_c$ are uncorrelated with covariance matrix $\mathbf{R}_c$ where

$$\mathbf{H}_c = \boldsymbol{\Lambda}_r^{1/2} \, \mathbf{U}_r^H \, \mathbf{H}_{\text{iid}} \, \mathbf{U}_t \, \boldsymbol{\Lambda}_t^{1/2} \overset{(a)}{\sim} \boldsymbol{\Lambda}_r^{1/2} \, \mathbf{H}_{\text{iid}} \, \boldsymbol{\Lambda}_t^{1/2} \tag{18}$$

$$\mathbf{R}_c = E\left[\mathbf{h}_c\mathbf{h}_c^H\right] = \boldsymbol{\Lambda}_t \otimes \boldsymbol{\Lambda}_r, \; \mathbf{h}_c \triangleq \text{vec}(\mathbf{H}_c). \tag{19}$$

The equality in (a) of (18) arises from the invariance of the distribution of $\mathbf{H}_{\text{iid}}$ under left and right unitary multiplications [1].

The spatial power matrix for the Kronecker model is denoted by $\mathbf{P}_k$. The Kronecker structure of $\mathbf{R}_c$ implies that the i-th column vector of $\mathbf{P}_k$ (denoted by $\mathbf{P}_{k,i}$) is

$$\mathbf{P}_{k,i} = \mathbf{\Lambda}_t[i] \cdot \left[\mathbf{\Lambda}_r[1], \mathbf{\Lambda}_r[2], \ldots, \mathbf{\Lambda}_r[N_r] \right]^T. \tag{20}$$

Note that this is a direct consequence of assuming separable statistics for $\mathbf{H}$ and does not hold in general. For the general case, separation of the transmit and receive domains can be artificially induced by using the *marginal* sum statistics (see Prop. 1). This fact highlights the limitations of the Kronecker model. The canonical framework results in a richer class of channels since it does not assume separability and an arbitrarily diagonal $\mathbf{R}_c$ is needed to model a general channel.

2) ULAs and the Virtual Representation: In [6], a virtual representation framework is proposed for systems with ULAs at both the transmitter and the receiver. In this case, $\mathbf{H}$ can be written as $\mathbf{A}_r \mathbf{H}_v \mathbf{A}_t^H$ where $\mathbf{A}_t$ and $\mathbf{A}_r$ are discrete Fourier transform (DFT) matrices. It is argued in [6] that the entries of $\mathbf{H}_v$ are approximately uncorrelated for finite number of antennas and the approximation becomes increasingly accurate as antenna dimensions increase. Thus, $\mathbf{A}_t$ and $\mathbf{A}_r$ serve as eigen-matrices in the virtual representation framework:

$$E\left[\mathbf{H}^H \mathbf{H}\right] = \mathbf{A}_t \mathbf{\Lambda}_{t,v} \mathbf{A}_t^H, \quad \mathbf{\Lambda}_{t,v} = E\left[\mathbf{H}_v^H \mathbf{H}_v\right], \tag{21}$$

$$E\left[\mathbf{H}\mathbf{H}^H\right] = \mathbf{A}_r \mathbf{\Lambda}_{r,v} \mathbf{A}_r^H, \quad \mathbf{\Lambda}_{r,v} = E\left[\mathbf{H}_v \mathbf{H}_v^H\right]. \tag{22}$$

Furthermore, Assumptions 1-4 made in the context of CM3 are satisfied by virtual representation. An important point to note is that while the transmit and the receive basis in CM3 are a function of the channel statistics and the entries of the canonical decomposition are *exactly* uncorrelated, the eigen-matrices $\mathbf{A}_t$ and $\mathbf{A}_r$ are fixed DFT matrices and entries of $\mathbf{H}_v$ are *approximately* uncorrelated. *Thus, in addition to the fact that the virtual representation for ULAs provides an intuitive physical interpretation where the eigenvectors $\mathbf{A}_t$ and $\mathbf{A}_r$ are beams in fixed virtual directions, it also makes transmit signal design easier since the transmit and the receive bases are fixed and do not change with the channel statistics.*

B. Transmit-Receive Eigen-spaces and Their Interaction

The decomposition of CM3 provides an equivalent representation in the eigen domain:

$$\mathbf{y}_c = \mathbf{H}_c \mathbf{x}_c + \mathbf{n}_c \tag{23}$$

where

$$\mathbf{y}_c \triangleq \mathbf{U}_r^H \mathbf{y}, \ \mathbf{x}_c \triangleq \mathbf{U}_t^H \mathbf{x}, \text{ and } \mathbf{n}_c \triangleq \mathbf{U}_r^H \mathbf{n}. \tag{24}$$

Thus, a linear transformation at the transmitter and the receiver results in $\mathbf{H}_c$ with independent entries. We note the following points.

- *Joint statistics* - $\mathbf{H}_c$ captures the *joint* transmitter-receiver statistics given by $\mathbf{P}_c$ which are in general non-separable, in comparison with the separable statistics of $\mathbf{P}_k$.
- *Degrees of freedom* - Define the DoF available in the channel as the entries of $\mathbf{H}_c$ having non-zero[4] variance. Thus,

$$\text{DoF} = \text{rank}(\mathbf{R}_c) = \text{rank}(\mathbf{R}) \leq N_t N_r. \tag{25}$$

  The i.i.d. channel has $\text{DoF} = N_t N_r$ and all the DoF have equal power. In correlated channels, however, the DoF is smaller and these DoF do not have equal power.
- *Parallel channels* - The parallel channels in the i.i.d. case have identical statistics and number $\min(N_t, N_r)$. In correlated channels, the non-zero columns of $\mathbf{P}_c$ expose the number of available parallel channels which is less than $\min(N_t, N_r)$, in general.

The last two observations signify the key differences in correlated versus i.i.d. channel modeling. Since the DoF and parallel channels are unequal, they should be excited appropriately for optimal transmission. *While the canonical model does not provide the same physical insight as the virtual representation (e.g., path partitioning), the mathematical similarities between the two models can be exploited.* This is witnessed by many recent works that explore the impact of independent entries in the case of virtual representation. See e.g., [16] for channel estimation; [7], [8], [19] and Sec. VI of this paper for capacity analysis; [37]–[39] for limited feedback system design; [40], [41] for non-coherent signal design; [42], [43] for space-time code design etc.

IV. STATISTICAL MODELS FOR MEASURED CHANNELS

In this work, we adopt the standard channel power normalization used in the MIMO literature: $\rho_c = N_t N_r$, where ρ_c is defined as

$$\rho_c \triangleq E[\text{Tr}(\mathbf{H}\mathbf{H}^H)] = E[\text{Tr}(\mathbf{H}_c\mathbf{H}_c^H)] = \text{Tr}(\mathbf{R}) = \text{Tr}(\mathbf{R}_c) = \sum_{ij} \mathbf{P}_c[i, j]. \tag{26}$$

A. Fitting Measured Channels with a Kronecker Model

Even though some initial studies [10]–[14] indicate that the Kronecker model is a good fit for 2×2 scenarios, further studies [18], [20]–[26] show that a non-separable modeling framework is more accurate. A non-separable framework in a Rayleigh fading setting is characterized by $N_t N_r$ statistical parameters, namely $\{\mathbf{P}_c[i, j]\}$. Initial difficulties on the tractability of the performance analysis of MIMO channels with such a general statistical description has led to the popularity of fitting the measured channel with a model characterized by fewer parameters. The following

---

[4]In practice, it is reasonable to define DoF as the number of entries in $\mathbf{P}_c$ that are larger than an *a priori*-determined threshold. The term "rank" in (25) should then be replaced with an appropriate definition of "effective rank."

proposition illustrates how a physical channel generated assuming non-separable statistics can be fitted with a Kronecker model.

Proposition 1: Consider a channel under CM3: $\mathbf{H} = \mathbf{U}_r \mathbf{H}_c \mathbf{U}_t^H$ with $\mathbf{P}_c[i,j] = E\left[|\mathbf{H}_c[i,j]|^2\right]$. A Kronecker fit for $\mathbf{H}$ is of the form $\mathbf{U}_r \mathbf{H}_k \mathbf{U}_t^H$ where $\mathbf{H}_k[i,j] \sim \mathcal{CN}(0, \mathbf{P}_k[i,j])$ with

$$\mathbf{P}_k[i,j] = \frac{\sum_l \mathbf{P}_c[i,l] \cdot \sum_k \mathbf{P}_c[k,j]}{\sum_{kl} \mathbf{P}_c[k,l]}. \tag{27}$$

Furthermore, the mapping in (27) *always* increases the DoF in the Kronecker fit for a scattering environment described by CM3.

Proof: Given a channel $\mathbf{H}$ that follows CM3, we attempt to fit a channel $\widetilde{\mathbf{H}}$ that follows the Kronecker model to it. From Sec. III-A, the general form of $\widetilde{\mathbf{H}}$ is $\mathbf{U}_{rk} \mathbf{\Lambda}_{rk}^{1/2} \mathbf{H}_{\text{iid}} \mathbf{\Lambda}_{tk}^{1/2} \mathbf{U}_{tk}^H$ for some appropriate choice of $\mathbf{U}_{tk}$, $\mathbf{U}_{rk}$, $\mathbf{\Lambda}_{tk}$ and $\mathbf{\Lambda}_{rk}$. By comparing the transmit and the receive covariance matrices with the two expansions, it can be checked that $\mathbf{U}_{tk} = \mathbf{U}_t$, $\mathbf{U}_{rk} = \mathbf{U}_r$ and $\mathbf{P}_k[i,j] \triangleq \mathbf{\Lambda}_{rk}[i] \, \mathbf{\Lambda}_{tk}[j]$ has to satisfy the relationship in (27). For the second part, note that

$$\mathbf{P}_k[i,j] \geq \frac{(\mathbf{P}_c[i,j])^2}{\sum_{kl} \mathbf{P}_c[k,l]} \tag{28}$$

and hence, $\mathbf{P}_k[i,j]$ is non-zero if $\mathbf{P}_c[i,j]$ is. Thus, the DoF in the Kronecker fit is always larger than the actual DoF with the canonical model. ∎

As an extreme artificial example of the above trend, consider a 4×4 system where the CM3 channel has $\mathrm{DoF} = 4$ and spatial power matrix $\mathbf{P}_c$ as in (29) below. It maps to a Kronecker model with $\mathrm{DoF} = 16$ and spatial power matrix $\mathbf{P}_k$ as below:

$$\mathbf{P}_c = 4 * \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{P}_k = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \tag{29}$$

In general, the Kronecker model spreads the degrees of freedom across the resulting $\mathbf{P}_k$ and thereby 'flattens' it since its statistics are based only on column and row sum statistics of the actual spatial power matrix. Note that while the transformation from $\mathbf{R}_c$ to $\mathbf{R}_k$ could lead to a change in rank, the transformation[5] from $\mathbf{H}_c$ to $\mathbf{H}_k$ does not.

B. Modeling Sparsity Mathematically

Another property suggested by fundamental electromagnetic studies [9], [29]–[35] as well as recent measurement campaigns [26, Figs. 9 and 11], [20], [21], [28] is that only a small subset

---

[5]As long as no row or column of $\mathbf{P}_c$ (and $\mathbf{P}_k$ following Prop. 1) are completely zero, the event where $\mathbf{H}_c$ (similarly for $\mathbf{H}_k$) is singular is a zero probability event [44].

of the $N_t N_r$ statistical parameters in CM3 are dominant enough to be leveraged towards reliable communications over practical SNR ranges. That is, measured wireless channels are *sparse*.

In this work, we compare the trends of the canonical and the Kronecker models across a large family of correlated/sparse channels. We provide two simple mathematical frameworks to generate large families of channel correlation information under CM3 and hence, from Prop. 1, under the Kronecker model. For this, we write $\mathbf{P}_c[i,j]$ as

$$\mathbf{P}_c[i,j] = N_t N_r \cdot \frac{p_{i,j}}{\sum_{ij} p_{i,j}} \tag{30}$$

where $\{p_{i,j}\}$ is a family of $N_t N_r$ random variables supported on $[0,1]$ that correspond to unnormalized[6] variances.

In sparse framework I, we set[7] $\{p_{i,j}\}$ to be i.i.d. with common mean and variance, μ and σ^2, respectively. A typical rich environment (intuitively, a 'near-i.i.d.' environment) is obtained by setting $\sigma^2 \approx 0$ with an i.i.d. channel corresponding to the extreme case of $\sigma^2 = 0$. As σ^2 increases, subject to the condition that $\sigma^2 \leq 1 - \mu^2$ (since $p_{i,j}$ are supported on $[0,1]$), $\{p_{i,j}\}$ get 'well-spread out' around μ. That is, there exists a large variability in the values of $\{\mathbf{P}_c[i,j]\}$, which intuitively reflects a correlated/sparse setting.

Despite a precise recipe for modeling in framework I, it could be difficult to systematically generate extremely sparse channels (where the fraction of dominant entries vanishes). In such settings, we propose sparse framework II in which we set $p_{i,j}$ as

$$p_{i,j} = q_{i,j} s_{i,j} \tag{31}$$

where $q_{i,j}$ is generated as described above (in framework I) and $s_{i,j}$ is an i.i.d. family of binary (0 or 1)- valued random variables with

$$\Pr(s_{i,j} = 1) = p = 1 - \Pr(s_{i,j} = 0). \tag{32}$$

Sparse channels can be generated systematically by adjusting the value of p appropriately. As p increases, the channel generated via (31) becomes *more richer* with the two frameworks coinciding for $p = 1$.

Note that frameworks I and II provide simple mathematical abstractions to model sparsity and their applicability in practice needs to be substantiated with further measurement studies. Nevertheless, as we will see, these simple models provide engineering intuition on the trends of capacity behavior.

V. CAPACITY OF CORRELATED MIMO CHANNELS

Towards this goal, we now briefly summarize some of the recent works on MIMO capacity. Prior to this summary, we state the channel state information (CSI) assumptions of this work.

---

[6]That is, $p_{i,j}$ have to be normalized, as in (30), to ensure that $\rho_c = N_t N_r$.

[7]The i.i.d. assumption on $\{p_{i,j}\}$ is made to simplify further analysis.

A. Channel State Information

We assume a coherent receiver architecture. That is, the receiver has perfect CSI. This is possible in practice by estimating the channel at the receiver using training symbols over a dedicated training period that lasts a significant portion of the channel coherence duration. We further assume that the statistics of the channel do not change over a reasonably long duration so that they can be acquired perfectly at the transmitter.

B. Ergodic Capacity

In this setting, the ergodic (or average) capacity at a transmit SNR of ρ is given by [2]

$$C_{\mathsf{erg}}(\rho) = \sup_{\mathbf{Q} \,:\, \mathbf{Q} \geq \mathbf{0}, \ \mathrm{Tr}(\mathbf{Q}) \leq \rho} E_{\mathbf{H}} \left[\log_2 \det \left(\mathbf{I} + \mathbf{H} \mathbf{Q} \mathbf{H}^H \right) \right] \tag{33}$$

where the optimization is over the set of trace-constrained, positive semi-definite matrices. While uniform-power (or full rank) signaling is optimal when no CSI is available at the transmitter, it is shown in [27], [45], [46] that the optimal $\mathbf{Q}$ to solve (33) has an eigen-decomposition

$$\mathbf{Q}_{\mathsf{opt}} = \mathbf{U}_t \boldsymbol{\Lambda}_{\mathsf{opt}} \mathbf{U}_t^H \tag{34}$$

where $\mathbf{U}_t$ is an eigen-matrix of $\mathbf{Q}_t = E\left[\mathbf{H}^H \mathbf{H} \right]$ and $\boldsymbol{\Lambda}_{\mathsf{opt}}$ is a positive semi-definite, diagonal matrix obtained via a numerical search. Closed-form solutions for $\boldsymbol{\Lambda}_{\mathsf{opt}}$ are not known; however, an iterative algorithm has been proposed in [27].

For any correlated channel, this algorithm converges to beamforming (or rank-1 signaling) in the asymptotically low-SNR regime and uniform-power signaling[8] in the asymptotically high-SNR regime. Thus, the low-SNR and the high-SNR ergodic capacities (denoted by $C_{\mathsf{low}}(\rho)$ and $C_{\mathsf{high}}(\rho)$, respectively) are given by

$$C_{\mathsf{low}}(\rho) \triangleq C_{\mathsf{erg}}(\rho) \Big|_{\mathbf{Q}_{\mathsf{opt}} \text{ as } \rho \to 0} = E\left[\log_2 \left(1 + \rho \sum_i \left| \mathbf{H}_c[i, j_{\max}] \right|^2 \right) \right] \tag{35}$$

$$C_{\mathsf{high}}(\rho) \triangleq C_{\mathsf{erg}}(\rho) \Big|_{\mathbf{Q}_{\mathsf{opt}} \text{ as } \rho \to \infty} = E\left[\log_2 \det \left(\mathbf{I}_{N_r} + \frac{\rho}{\mathsf{rank}(\mathbf{P}_c)} \mathbf{H}_c \mathbf{H}_c^H \right) \right] \tag{36}$$

where $j_{\max} = \arg\max_j \sum_i \mathbf{P}_c[i, j]$ corresponds to the dominant transmit eigen-direction. In general, at an intermediate SNR, the optimal rank of $\mathbf{Q}$ is non-decreasing (as ρ increases) with precise estimates available for the transient-SNR's when a particular rank signaling scheme becomes optimal [47].

---

[8]Without any loss in generality, we assume that no column of $\mathbf{P}_c$ has all zero entries.

C. Outage Capacity

It is also well-understood that the ergodic capacity is an insufficient metric to understand the fundamental impact fading has on achievable data rates and the notion of outage capacity [2], [48] at an outage probability of $q\%$ is relevant. The outage capacity is the maximum rate that is guaranteed for at least $(100 - q)\%$ of the channel realizations and is defined as

$$C_{\mathsf{out},\,q}(\rho) \triangleq \sup_{R \geq 0} (R) \quad \text{s.t.} \quad \Pr\left(\log_2 \det\left[\mathbf{I} + \mathbf{HQH}^H\right] < R\right) \leq \frac{q}{100}. \tag{37}$$

Many recent works have shown that Gaussian approximations to $C_{\mathsf{out},\,q}(\rho)$ with mean and variance given by $C_{\mathsf{erg}}(\rho)$ and $V(\rho)$, the variance of capacity, are accurate in the large-system limit; see [36] and references therein. Thus, in the large-system limit, $C_{\mathsf{out},\,q}(\rho)$ can be efficiently approximated as

$$C_{\mathsf{out},\,q}(\rho) = C_{\mathsf{erg}}(\rho) - x_q \sqrt{V(\rho)} + o(1) \tag{38}$$

where x_q is the unique solution to $\mathrm{erfc}\left(x_q/\sqrt{2}\right) = 2q$ with $\mathrm{erfc}(\cdot)$ denoting the complementary error function.

VI. COMPARATIVE STUDY OF CAPACITY OF KRONECKER AND CANONICAL MODELS

An important point to note from (38) is that the outage capacity is determined upon knowledge of $C_{\mathsf{erg}}(\rho)$ and $V(\rho)$. The main focus of this section is thus on understanding $C_{\mathsf{erg}}(\rho)$ and $V(\rho)$ when a MIMO channel (with non-separable statistics) is fitted with a Kronecker model as in Prop. 1. We denote by $C_{\mathsf{erg,\,can}}(\rho)$ and $C_{\mathsf{erg,\,kron}}(\rho)$, the means of capacity under the two models and by $V_{\mathsf{can}}(\rho)$ and $V_{\mathsf{kron}}(\rho)$, the variances of capacity under these models.

In this section, we provide good estimates for the above quantities under certain conditions. While an analytical understanding of these quantities for all SNRs seems difficult, it is possible to obtain engineering intuition by studying the mismatches (between the two capacities) at the low- and the high-SNR extremes under a large-system assumption. Since the convergence to the large-system regime is typically fast (see e.g., [36] and references therein which point out that good agreement is possible even with 4 or 8 antennas) we expect this analysis to be useful in making meaningful conclusions in the finite antenna regime.

A. Low-SNR Extreme

As noted in Sec. V, beamforming to the statistically dominant[9] transmit eigen-mode (which is the same irrespective of whether beamforming is done based on the statistics of $\mathbf{H}_c$ or $\mathbf{H}_k$) is optimal from an ergodic capacity perspective in the low-SNR regime. However, many works in

---

[9]Without loss in generality, let all the N_r entries in the dominant column $\{\mathbf{P}_c[i, j_{\max}], i = 1, \cdots, N_r\}$ be non-zero.

the literature define the low-SNR regime *imprecisely* as "$\rho \to 0$." It is useful to define a transient-SNR, ρ_{low}, such that beamforming is capacity-optimal if $\rho < \rho_{\text{low}}$. Some works (see [8], [49] and references therein) define ρ_{low} *implicitly* in terms of means of certain random variables that are related to $\mathbf{P}_c$, but are nevertheless difficult to compute in closed-form. In [47], using tools[10] from random matrix theory, it is shown that

$$\rho_{\text{low}} \approx \frac{1}{\sum_{i=1}^{N_r} \mathbf{P}_c[i, j_{\max}]}. \tag{39}$$

Capacity Computation: We first develop a general low-SNR characterization of MIMO capacity in the canonical case, and then leverage this result to the Kronecker case. For this, we define[11] $\widehat{\rho}_{\text{low, can}}$:

$$\widehat{\rho}_{\text{low, can}} \triangleq \frac{1}{\sum_{i=1}^{N_r} \mathbf{P}_c[i, j_{\max}]} \cdot \frac{1}{\gamma_0}, \quad \gamma_0 = 1 + \frac{\sqrt{N_r \sum_{i=1}^{N_r} (\mathbf{P}_c[i, j_{\max}])^2}}{2 \sum_{i=1}^{N_r} \mathbf{P}_c[i, j_{\max}]}. \tag{40}$$

The importance of $\widehat{\rho}_{\text{low, can}}$ is that $I(\rho)$, the average mutual information with statistical beamforming, is given by

$$I(\rho) = \log_2(e) \cdot \rho \cdot \sum_{i=1}^{N_r} \mathbf{P}_c[i, j_{\max}] \cdot (1 + o(1)), \quad \rho < \widehat{\rho}_{\text{low, can}}. \tag{41}$$

It should be noted that $C_{\text{low}}(\rho)$ shows the same trends as $I(\rho)$. This is the content of the following theorem.

Theorem 1: There exist positive constants $c, \ell > 0, m > 1/2$ (all independent of $\mathbf{P}_c, N_t$ and N_r) such that

$$\log_2(e) \cdot \delta \left(1 - \frac{2^\ell \sqrt{\kappa_c}}{N_r^m} - \frac{\kappa_c \delta}{\gamma_0} \right) \leq C_{\text{erg, can}}(\rho) \leq \log_2(e) \cdot \delta$$

$$\text{for all} \quad \rho = \frac{\delta}{\sum_{i=1}^{N_r} \mathbf{P}_c[i, j_{\max}]}, \ 0 < \delta < c, \tag{42}$$

where κ_c is defined as

$$\kappa_c \triangleq 1 + \frac{\sum_{i=1}^{N_r} (\mathbf{P}_c[i, j_{\max}])^2}{\left(\sum_{i=1}^{N_r} \mathbf{P}_c[i, j_{\max}] \right)^2} \tag{43}$$

and γ_0 is as in (40). Alternately, the above statement can be recast as

$$C_{\text{erg, can}}(\rho) = \log_2(e) \cdot \rho \cdot \sum_i \mathbf{P}_c[i, j_{\max}] (1 + o(1)) \tag{44}$$

---

[10]Also, see [50] which points this out from a reconfigurable antennas point-of-view.

[11]The difference between $\rho_{\text{low,can}}$ and $\widehat{\rho}_{\text{low,can}}$ is that while beamforming is *exactly* capacity-optimal below $\rho_{\text{low,can}}$, it is only near-optimal below $\widehat{\rho}_{\text{low,can}}$. Nevertheless, note that if $\mathbf{P}_c[i, j_{\max}] = 1$ for all i, $\widehat{\rho}_{\text{low,can}}$ reduces to $\frac{2}{3N_r}$ and thus the trends of $\widehat{\rho}_{\text{low,can}}$ are similar to that of $\rho_{\text{low,can}}$.

with the $o(1)$ factor converging to 0 as $N_r \to \infty$ and $\delta \to 0$.

Proof: See Appendix B. ∎

From (27), it follows that if $\{\mathbf{P}_c[i, j_{\max}]\}$ are non-zero, so are $\{\mathbf{P}_k[i, j_{\max}]\}$. It is thus easy to specialize Theorem 1 to the Kronecker case (associated with $\widehat{\rho}_{\text{low, kron}}$) and compare the two results.

Capacity Comparison:

Theorem 2: Let the low-SNR regime be defined as $\rho < \widehat{\rho}_{\text{low}}$ where

$$\widehat{\rho}_{\text{low}} \triangleq \min\left(\widehat{\rho}_{\text{low, can}}, \widehat{\rho}_{\text{low, kron}}\right). \tag{45}$$

In this regime, the following conclusions hold for the dominant terms of the capacity quantities.

- (a) The dominant terms of the ergodic capacity under the two models is the same. In particular, we have

$$C_{\text{erg, can}}(\rho) = C_{\text{erg, kron}}(\rho) = \log_2(e)\rho \cdot \sum_i \mathbf{P}_c[i, j_{\max}]. \tag{46}$$

- (b) The dominant terms of the variances satisfy

$$\frac{V_{\text{can}}(\rho)}{(\log_2(e)\rho)^2} = \sum_i \left(\mathbf{P}_c[i, j_{\max}]\right)^2, \qquad \frac{V_{\text{kron}}(\rho)}{(\log_2(e)\rho)^2} = \sum_i \left(\mathbf{P}_k[i, j_{\max}]\right)^2. \tag{47}$$

- (c) Let $\mathbf{P}_c$ be row-permuted such that $\{\sum_{k=1}^{N_t} \mathbf{P}_c[i, k], \ i = 1, \cdots, N_r\}$ is arranged in decreasing order. Further, if the entries of $\mathbf{P}_c$ satisfy

$$\frac{\sum_{k \neq j_{\max}} \mathbf{P}_c[i, k]}{\sum_{k \neq j_{\max}} \mathbf{P}_c[i+1, k]} \leq \frac{\mathbf{P}_c[i, j_{\max}]}{\mathbf{P}_c[i+1, j_{\max}]} \text{ for all } 1 \leq i \leq N_r - 1, \tag{48}$$

then $V_{\text{can}}(\rho) \geq V_{\text{kron}}(\rho)$ as $\rho \to 0$.

Proof: See Appendix C. ∎

The condition in (48) implies that the fraction of power captured in the beamforming direction by a receiver decreases in the same order as the *total* power captured by the receivers. For example, in the case of regular channels (see Footnote 3), it is easy to check that (48) holds trivially. In fact, for regular channels, it can be checked that

$$\frac{V_{\text{can}}(\rho)}{V_{\text{kron}}(\rho)} = \frac{N_r \sum_{i=1}^{N_r} \left(\mathbf{P}_c[i, j_{\max}]\right)^2}{\left(\sum_{i=1}^{N_r} \mathbf{P}_c[i, j_{\max}]\right)^2} \geq 1 \tag{49}$$

due to the Cauchy-Schwarz inequality. It also seems like the condition in (48) is necessary to ensure that $V_{\text{can}}(\rho) \geq V_{\text{kron}}(\rho)$. For example, it can be checked that $V_{\text{can}}(\rho) < V_{\text{kron}}(\rho)$ in the following 2×2 case where

$$\mathbf{P}_c = \begin{bmatrix} 1 & A(1+\epsilon) \\ 1 & A \end{bmatrix}, \ A \leq \frac{2}{2+\epsilon}, \ \epsilon > 0 \tag{50}$$

and (48) does not hold. Nevertheless, in the large-system regime, we have the following con-
clusions for the probabilistic sparse frameworks I and II, introduced in Sec. IV-B.

Proposition 2: First, recall that I is a special case of II with $p = 1$.

- (a) The probability with which the condition in (48) holds converges to 1 as $\{N_t, N_r\} \to \infty$.
  Hence, $V_{\text{can}}(\rho) \geq V_{\text{kron}}(\rho)$ for "almost all"[12] sparse scattering environments generated from
  either framework.

- (b) In particular, if $0 < m \leq q_{i,j} \leq M$ with $E[q_{i,j}] = \mu$ and $\text{Var}(q_{i,j}) = \sigma^2$, we have

$$1 \leq \frac{V_{\text{can}}(\rho)}{V_{\text{kron}}(\rho)} \leq \frac{1}{p} \cdot \frac{(M + m)^2}{4Mm}. \tag{51}$$

  More specifically, we have $\frac{V_{\text{can}}(\rho)}{V_{\text{kron}}(\rho)} \to \frac{1}{p}\left(1 + \frac{\sigma^2}{\mu^2}\right)$.

- (c) Equality in the lower bound of (51) is achieved when $\mathbf{H}$ is i.i.d. If $N_t = N_r = N$ such
  that $\frac{MN}{M+m}$ and $\frac{mN}{M+m}$ are integers, equality in the upper bound is approached as $N \to \infty$ by

$$\mathbf{P}_c^T = \frac{N}{M+m} \cdot \begin{bmatrix} M+m & \underbrace{m \ \cdots \ m}_{\frac{MN}{M+m}-1} & \underbrace{M \ \cdots \ M}_{\frac{mN}{M+m}} \\ 0 & M & & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & \ddots & \vdots \\ 0 & & & M & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & m & \\ \vdots & \vdots & \ddots & \vdots & & \ddots \\ 0 & 0 & \cdots & 0 & & & m \end{bmatrix}. \tag{52}$$

Proof: See Appendix D. ∎

Note that the channel corresponding to $\mathbf{P}_c$ in (52) is such that $\boldsymbol{\Sigma}_t$ has at least $N\left(1 - \frac{m}{M}\right)$
dominant eigenvalues whereas the eigenvalues of $\boldsymbol{\Sigma}_r$ are all equal to $M + m$. It is surprising that
channels that are 'near-well-conditioned' on both the transmitter and the receiver sides ($\mathbf{H}_{\text{iid}}$ and
the channel in (52)) could either maximize or minimize $\frac{V_{\text{can}}(\rho)}{V_{\text{kron}}(\rho)}$ depending on the distribution of
non-zero entries in $\mathbf{P}_c$.

Discussion: The above results show that the ergodic capacities remain the same under the
canonical and the Kronecker models for all channels in the low-SNR regime. Thus, the dominant
factors in understanding outage capacity (rate vs. reliability trade-off) in (38) are the variances of
capacity. Since $V_{\text{can}}(\rho) \geq V_{\text{kron}}(\rho)$ for almost all sparse channels, the outage capacity under the
Kronecker model is always steeper than the outage capacity under the canonical model (except

---

[12]Technically, this statement has to be read as: "with probability 1 on the probability space corresponding to $\{p_{i,j}\}$."
Henceforth, we will not bother with this detail.

for i.i.d. $\mathbf{H}_c$ where they are equally steep). Furthermore, the differential in steepness increases as the channel becomes more sparse.

In other words, at high levels of operational reliability, the Kronecker model overestimates capacity while it switches roles and underestimates capacity at low levels of reliability. However, the smallness of the capacity values generally means that these trends are not prominent when we plot outage capacity in the low-SNR regime. For example, Figs. 2-4 plot the cumulative distribution function (CDF) of capacity (at $-10, 10$ and 30 dB SNRs) for three 8×8 channels generated[13] to portray: i) A typical sparse setting, ii) A setting with intermediate level of richness, and iii) A typical rich setting. The spatial power matrices are given by

$$\mathbf{P}_{c,\,\text{sparse}} = \begin{bmatrix} 43.2693 & 2.5367 & 0.4362 & 0.5569 & 0.1004 & 0.1805 & 0.1413 & 0.2009 \\ 3.1184 & 1.9519 & 2.4028 & 1.4193 & 0.1007 & 0.1635 & 0.2420 & 0.1864 \\ 0.4997 & 0.6135 & 0.5352 & 0.5998 & 0.0502 & 0.4960 & 0.1892 & 0.1532 \\ 0.0574 & 0.0671 & 0.3255 & 0.2233 & 0.0653 & 0.1082 & 0.1258 & 0.0695 \\ 0.0702 & 0.1138 & 0.1534 & 0.0767 & 0.0327 & 0.0482 & 0.1087 & 0.0426 \\ 0.0192 & 0.0564 & 0.0540 & 0.1060 & 0.0908 & 0.0608 & 0.0512 & 0.0426 \\ 0.0147 & 0.0174 & 0.0090 & 0.0132 & 0.0257 & 0.0266 & 0.0287 & 0.0181 \\ 0.0037 & 0.0032 & 0.0026 & 0.0034 & 0.0029 & 0.0050 & 0.0038 & 0.0036 \end{bmatrix} \tag{53}$$

while for rich scattering, it is

$$\mathbf{P}_{c,\,\text{rich}} = \begin{bmatrix} 8.6986 & 3.7188 & 0.7361 & 1.1644 & 2.1912 & 1.4959 & 1.4674 & 1.6262 \\ 2.5481 & 3.5585 & 0.5683 & 0.7660 & 1.4141 & 1.3985 & 1.3766 & 1.2021 \\ 0.9910 & 2.1553 & 0.2217 & 0.3105 & 3.0150 & 0.6737 & 0.6415 & 1.7003 \\ 0.0793 & 0.1347 & 0.0685 & 0.0750 & 0.1279 & 0.1439 & 0.1218 & 0.0778 \\ 0.1651 & 0.1946 & 0.0952 & 0.2151 & 0.3063 & 0.2179 & 0.2742 & 0.2119 \\ 1.2612 & 0.9009 & 0.3837 & 0.6108 & 1.3185 & 0.7591 & 0.9227 & 1.7308 \\ 0.3696 & 1.0783 & 0.4618 & 0.7504 & 0.5129 & 0.5620 & 1.4507 & 0.7081 \\ 0.7332 & 0.2221 & 0.1895 & 0.3626 & 0.6765 & 0.4681 & 0.6841 & 0.7337 \end{bmatrix}, \tag{54}$$

and $\mathbf{P}_{c,\,\text{intermediate}} = \frac{\mathbf{P}_{c,\,\text{sparse}} + \mathbf{P}_{c,\,\text{rich}}}{2}$.

Note that $\sum_{ij} \mathbf{P}_{c,\,\bullet} = 64$ for all the three channels and the ratio of the largest to the smallest transmit eigenvalue decreases from 100.4 to 9.31 and 5.45 as the channel becomes progressively richer. The ratio of the largest to the smallest receive eigenvalue decreases from 1682 to 36.6 and 25.5 as the channel becomes richer. The channel realizations are generated as

$$\mathbf{H}_{\bullet} = \mathbf{H}_{\text{iid}} \odot (\mathbf{P}_{c,\,\bullet})^{1/2} \tag{55}$$

---

[13]The spatial power matrices for this experiment have been generated artificially to mimic certain typical scattering environments, and not using the sparse frameworks of Sec. IV-B.
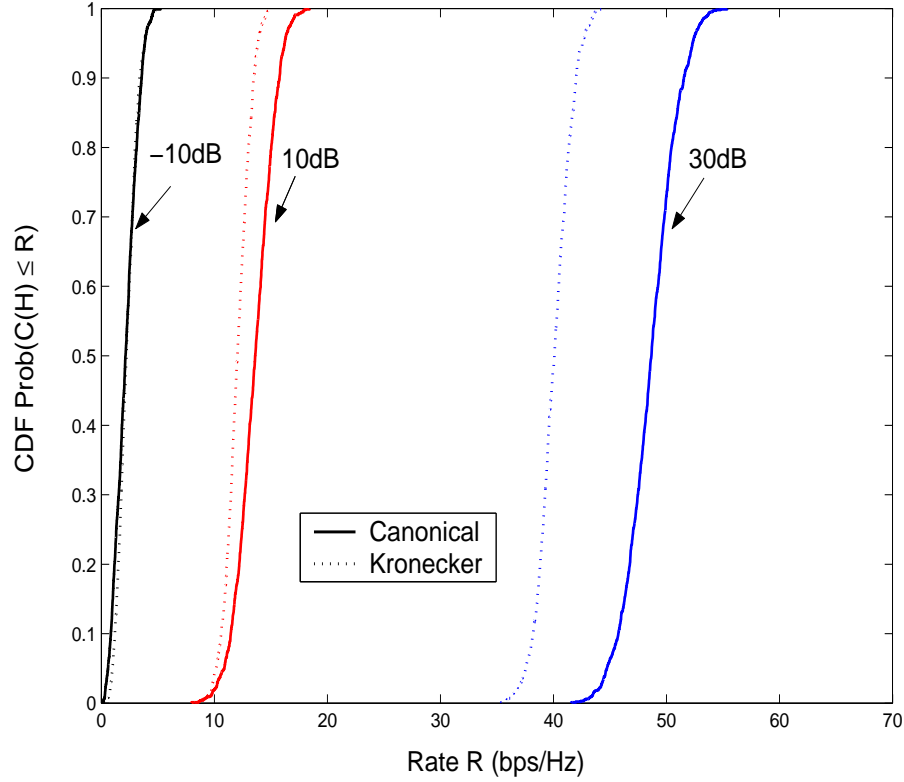
Fig. 2. Capacity CDFs of a *sparse* channel with canonical and Kronecker models at $-10, 10$ and 30 dB SNRs.

where $\mathbf{H}_{\text{iid}}$ is an i.i.d. channel and $(\mathbf{P}_{c,\bullet})^{1/2}$ is the element-wise square-root of the spatial power matrix.

Spectral Efficiency: Another characterization of low-SNR performance is in the context of spectral efficiency [51] (equivalently, $C_{\text{erg},\bullet}(\rho)$ vs. ρ behavior). We now present the connections between the canonical and the Kronecker models to the two key figures-of-merit in low-SNR communications: i) Minimum energy per bit necessary for reliable communication, $\frac{E_b}{N_{\text{o min}}}$, and ii) Wideband slope, S_0. For a multi-antenna channel, these two metrics are given by [51]

$$\frac{E_b}{N_{\text{o min}}} = \frac{\log_e(2)}{E\left[\text{Tr}\left(\mathbf{HQH}^H\right)\right]}, \text{ and } S_0 = 2 \cdot \frac{\left(E\left[\text{Tr}\left(\mathbf{HQH}^H\right)\right]\right)^2}{E\left[\text{Tr}\left((\mathbf{HQH}^H)^2\right)\right]} \tag{56}$$

where the input covariance matrix, $\mathbf{Q} = \text{diag}\left(\mathbf{Q}[i]\right)$, is low-SNR capacity-achieving and unit trace constrained.

When there is *only* one dominant transmit eigen-mode, beamforming to this mode is spectral efficiency-optimal. If there are r dominant eigen-modes with $r > 1$, any $\mathbf{Q}$ that excites any of the r modes with any weightage is ergodic capacity-optimal. However, [51] points out that uniform-power signaling over these r modes is necessary to maximize spectral efficiency. We consider these two cases separately in the following theorem.
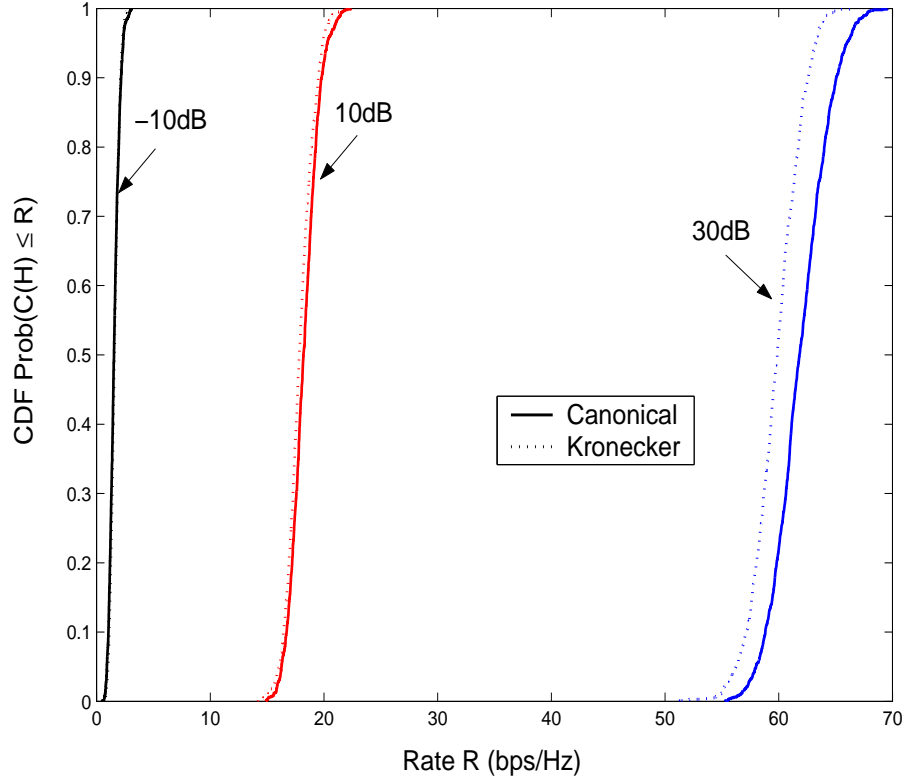
Fig. 3. Capacity CDFs of a channel that has an intermediate level of richness with canonical and Kronecker models at $-10, 10$ and 30 dB SNRs.

Theorem 3: If $r = 1$, the minimum energies per bit are given by

$$\frac{E_b}{N_{\text{o min, can}}} = \frac{E_b}{N_{\text{o min, kron}}} = \frac{\log_e(2)}{\sum_i \mathbf{P}_c[i, j_{\max}]} \overset{(a)}{\to} \frac{\log_e(2)}{N_r p \mu} \tag{57}$$

where the convergence in (a) is for the sparse framework II. An application of the Gaussian moment factoring theorem [52] with the optimal input shows that

$$S_{0, \text{can}} = 2 \cdot \frac{\left(\sum_i \mathbf{P}_c[i, j_{\max}]\right)^2}{\sum_i \left(\mathbf{P}_c[i, j_{\max}]\right)^2 + \left(\sum_i \mathbf{P}_c[i, j_{\max}]\right)^2}, \tag{58}$$

$$S_{0, \text{kron}} = 2 \cdot \frac{\left(\sum_i \mathbf{P}_k[i, j_{\max}]\right)^2}{\sum_i \left(\mathbf{P}_k[i, j_{\max}]\right)^2 + \left(\sum_i \mathbf{P}_k[i, j_{\max}]\right)^2}. \tag{59}$$

With framework II, we have

$$S_{0, \text{can}} \to \frac{2 N_r \mu^2 p}{(N_r p + 1)\mu^2 + \sigma^2}, \quad \text{and} \quad S_{0, \text{kron}} \to \frac{2 N_r}{N_r + 1}. \tag{60}$$

If $r > 1$, the energies per bit are the same as in (57). The wideband slopes generalize to

$$S_{0, \text{can}} \to \frac{2 N_r r \mu^2 p}{\mu^2(N_r p + 1 + (r-1)p) + \sigma^2}, \quad \text{and} \quad S_{0, \text{kron}} \to \frac{2 N_r r}{N_r + r}. \tag{61}$$
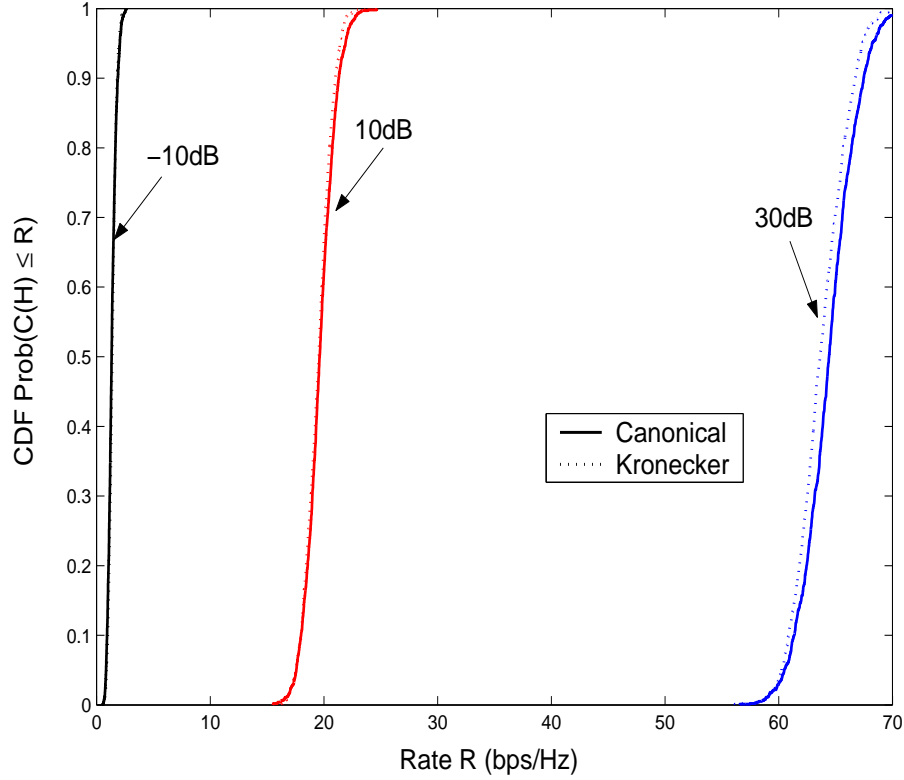
Fig. 4. Capacity CDFs of a *rich* channel with canonical and Kronecker models at $-10, 10$ and 30 dB SNRs.

Proof: With $r = 1$, the conclusion about energy per bit is straightforward. The expression for the wideband slope follows immediately from the fact proved in Appendix C:

$$N_r p^2 \mu^2 \leftarrow \sum_i \left(\mathbf{P}_k[i, j_{\max}]\right)^2 < \sum_i \left(\mathbf{P}_c[i, j_{\max}]\right)^2 \rightarrow N_r p \left(\mu^2 + \sigma^2\right). \tag{62}$$

For $r > 1$, see Appendix E. ∎

It can be checked that $S_{0,\text{can}} < S_{0,\text{kron}}$ in either case. However, this conclusion is not *easily* reflected in Figs. 2-4 due to two reasons:

- $\frac{E_b}{N_{\text{o}} \min}$, which is the same for both the channel models (in both $r = 1$ and $r > 1$ cases), is the most important figure of merit at low-SNR and corresponds to first order variation in ergodic capacity with SNR while S_0 corresponds to second order variation at low-SNR.
- The discrepancy in S_0 for the two models is small. In fact, we have

$$|S_{0,\text{kron}} - S_{0,\text{can}}| \leq \frac{\sum_i \left(\left(\mathbf{P}_c[i, j_{\max}]\right)^2 - \left(\mathbf{P}_k[i, j_{\max}]\right)^2\right)}{\left(\sum_i \mathbf{P}_c[i, j_{\max}]\right)^2} = \frac{\sigma^2}{p\mu^2} \cdot \mathcal{O}\left(\frac{1}{N_r}\right) \tag{63}$$

for the $r = 1$ case, and

$$|S_{0,\text{kron}} - S_{0,\text{can}}| \leq \frac{\sigma^2}{p\mu^2} \cdot \mathcal{O}\left(\frac{N_r r}{(N_r + r)^2}\right) \tag{64}$$

for the $r > 1$ case. In the second case, the difference in wideband slopes is $\mathcal{O}\left(\frac{1}{N_r}\right)$ for finite values of r.

B. High-SNR Extreme

We now make two assumptions on the random matrix channel $\mathbf{H}_c$ to aid[14] in capacity analysis: 1) $N_t = N_r = N$, and 2) $\mathsf{rank}(\mathbf{H}_c) = N$ a.s. Note that the second condition is equivalent to assuming that none of $\{\sum_i \mathbf{P}_c[i,j]\}$ and $\{\sum_j \mathbf{P}_c[i,j]\}$ are zero. From the discussion following Prop. 1, we also have $\mathsf{rank}(\mathbf{H}_k) = N$ a.s..

In this setting, the capacity random variables under the two models are given by

$$C_{\mathsf{can}}(\rho, \mathbf{H}) \triangleq \log_2 \det \left(\mathbf{I}_N + \frac{\rho}{N}\mathbf{H}_c\mathbf{H}_c^H\right) \overset{(a)}{=} \log_2 \det \left(\mathbf{H}_c\mathbf{H}_c^H\right) + N \log_2 \left(\frac{\rho}{N}\right) + \mathcal{O}\left(\frac{1}{\rho}\right) \quad (65)$$

$$C_{\mathsf{kron}}(\rho, \mathbf{H}) \triangleq \log_2 \det \left(\mathbf{I}_N + \frac{\rho}{N}\mathbf{H}_k\mathbf{H}_k^H\right) \overset{(b)}{=} \log_2 \det \left(\mathbf{H}_k\mathbf{H}_k^H\right) + N \log_2 \left(\frac{\rho}{N}\right) + \mathcal{O}\left(\frac{1}{\rho}\right) \quad (66)$$

where in (a) we have used both Assumptions 1) and 2), and in (b), we have used the fact that $\mathsf{rank}\left(\mathbf{H}_k\right) = N = \mathsf{rank}\left(\mathbf{H}_c\right)$ a.s. Hence the statistics of $C_{\mathsf{can}}(\rho, \mathbf{H})$ and $C_{\mathsf{kron}}(\rho, \mathbf{H})$ at high-SNR are related to the moments of $\log_2 \det \left(\mathbf{H}_c\mathbf{H}_c^H\right)$ and $\log_2 \det \left(\mathbf{H}_k\mathbf{H}_k^H\right)$, respectively. We now perform a large-system analysis of these random log-determinants.

Stochastic Approximation for the Canonical Case: In the case of $\mathbf{H}_{\mathsf{iid}}$ ($\mathbf{P}_c[i,j] = 1$ for all i, j), this analysis is simplified by what is known as the *Bartlett decomposition (or bidiagonalization) of a sample covariance matrix* [44], [53], [54]. The decomposition states that there exist independent random variables $\mathbf{Z}_i$ on some probability space such that

$$\mathbf{Z} \triangleq \det \left(\mathbf{H}_{\mathsf{iid}} \mathbf{H}_{\mathsf{iid}}^H\right) \sim \prod_{i=1}^{N} \mathbf{Z}_i, \ \mathbf{Z}_i \sim \sum_{j=i}^{N} |\mathbf{H}_{\mathsf{iid}}[i,j]|^2 \sim \frac{1}{2} \chi^2 \left(2(N-i+1)\right) \quad (67)$$

where $\chi^2(2k)$ is a central chi-squared random variable with $2k$ degrees of freedom.

On the other hand, computing $\log_2 \det \left(\mathbf{H}_c\mathbf{H}_c^H\right)$ in closed-form is extremely difficult because $\{\mathbf{P}_c[i,j]\}$, in general, possess no structure and a Bartlett-type decomposition for $\det \left(\mathbf{H}_c\mathbf{H}_c^H\right)$ is not known. Nevertheless, a tight stochastic approximation for $C_{\mathsf{erg, can}}(\rho)$ is still possible and for this, we need the following notation from [55].

We say that a random variable $\mathbf{X}_2$ upper bounds a random variable $\mathbf{X}_1$ (and denote it by $\mathbf{X}_1 \lesssim \mathbf{X}_2$) if

$$\Pr\left(\mathbf{X}_1 < x\right) \ \geq \ \Pr\left(\mathbf{X}_2 < x\right) \text{ for all } x \in \mathbb{R}. \quad (68)$$

---

[14]The first condition can be relaxed with some advanced random matrix theory techniques that are out-of-scope here. If this is done and we obtain results for arbitrary N_t and N_r, then the second condition can be assumed without any loss in generality since we can always ignore those columns/rows with zero power. Nevertheless, for simplicity of analysis, we assume both conditions.

The following lemma provides a statistical "bound" and a useful stochastic approximation for $\det\left(\mathbf{H}_c \mathbf{H}_c^H\right)$.

Lemma 5 (Girko): Let $\widetilde{\mathbf{H}}[i, j]$ be independent and distributed as $\mathcal{CN}(0, p_{i,j})$. Then,

$$\mathbf{Z} \cdot \prod_{i=1}^{N} \min_{j} p_{i,j} \;\lesssim\; \det\left(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H\right) \;\lesssim\; \mathbf{Z} \cdot \prod_{i=1}^{N} \max_{j} p_{i,j} \tag{69}$$

where $\mathbf{Z}$ is as in (67). Moreover, there exist independent random variables $\widetilde{\mathbf{Z}}_i$, $i = 1 \cdots N$ on some probability space such that $\det\left(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H\right)$ can be well-approximated as

$$\det\left(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H\right) \approx \prod_{i=1}^{N} \widetilde{\mathbf{Z}}_i, \; \widetilde{\mathbf{Z}}_i \sim i \; \frac{\sum_{j=1}^{N} \left|\widetilde{\mathbf{H}}[i, j]\right|^2}{N}. \tag{70}$$
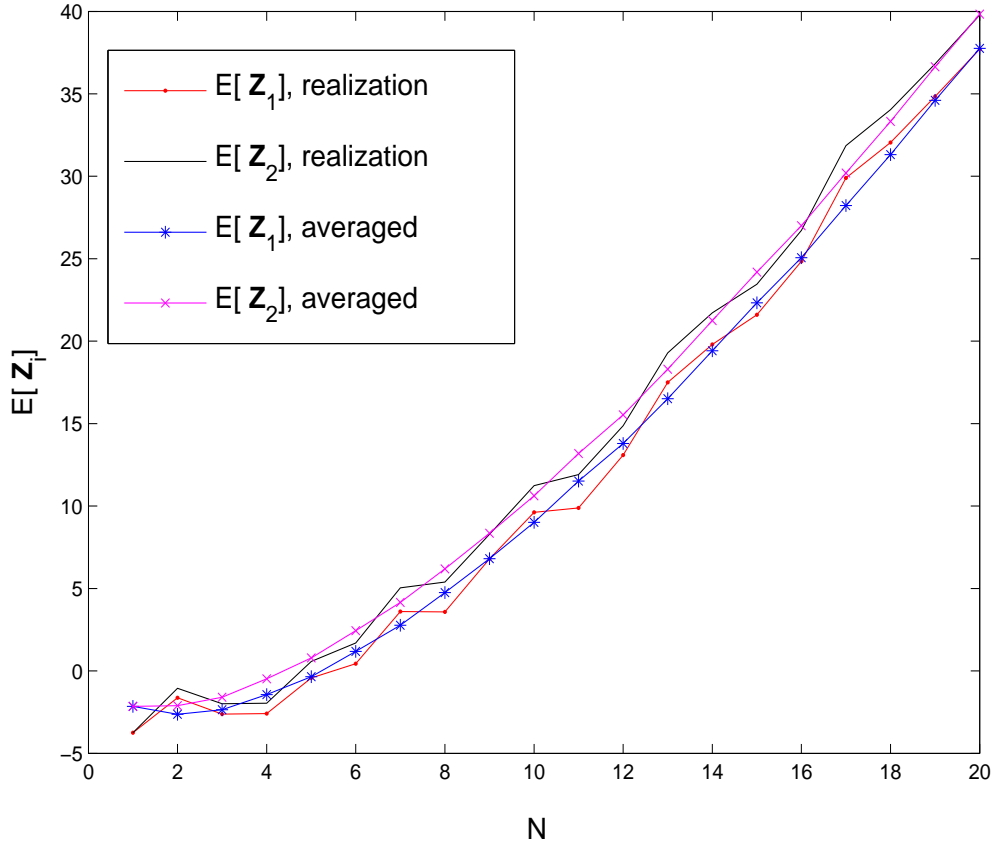
Proof: See Appendix F. ∎

Fig. 5. Comparison of means of $\mathbf{Z}_i$ (as a function of N) for a typical scattering environment and averaged over many scattering environments.

Numerical studies indicate that the approximation in Lemma 5 is close for a large class of random matrices *even* for small values of N. Furthermore, this approximation gets more accurate

as N increases for a large class of random matrices. This fact is illustrated in Fig. 5 where we plot $E[\mathbf{Z}_1]$ and $E[\mathbf{Z}_2]$ as a function of matrix dimension N with

$$\mathbf{Z}_1 = \log_2 \det(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H) \text{ and } \mathbf{Z}_2 = \sum_{i=1}^{N} \log_2 \left(\frac{i \sum_{j=1}^{N} |\widetilde{\mathbf{H}}[i,j]|^2}{N} \right). \tag{71}$$

The first set corresponds to a typical scattering environment where $\{p_{i,j}\}$ are chosen i.i.d. from a uniform distribution on $[0,1]$ (in particular, $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{12}$). The second set corresponds to a smoothed version of the first set where we also average over many different scattering environments. Here, we have averaged over 5000 independent scattering environments and the plot shows that the approximation is very accurate on average. In the rest of the paper, we assume that the approximation in (70) is accurate. Nevertheless, its rigorous use is contingent on further studies that have to establish its preciseness. This will be the subject of future work.

Capacity Computation and Comparison:

Theorem 4: With the sparse frameworks of Sec. IV-B, good estimates can be obtained for ergodic capacity in the high-SNR extreme.

- (a) The ergodic capacity under the Kronecker model converges to

$$C_{\text{erg, kron}}(\rho) \;\rightarrow\; N \log_2 \left(\frac{\rho N^2}{\sum_{ij} p_{i,j}} \right) + \sum_{i=1}^{N} \log_2 \left(\frac{i}{N} \right) + K_{\text{kron}} + \mathcal{O}\left(\frac{1}{\rho} \right) \tag{72}$$

$$K_{\text{kron}} \;=\; \sum_{i=1}^{N} \log_2 \left(\frac{\sum_l p_{i,l} \sum_k p_{k,i}}{\sum_{kl} p_{k,l}} \right) \tag{73}$$

  whereas under the canonical model, it is well-approximated (with the approximation approaching an equality as $N \to \infty$ following the previous discussion) by

$$C_{\text{erg, can}}(\rho) \;\approx\; N \log_2 \left(\frac{\rho N^2}{\sum_{ij} p_{i,j}} \right) + \sum_{i=1}^{N} \log_2 \left(\frac{i}{N} \right) + K_{\text{can}} + \mathcal{O}\left(\frac{1}{\rho} \right) \tag{74}$$

$$K_{\text{can}} \;=\; \frac{1}{2} \left\{ \sum_{i=1}^{N} \log_2 \left(\frac{\sum_{j=1}^{N} p_{i,j}}{N} \right) + \sum_{j=1}^{N} \log_2 \left(\frac{\sum_{i=1}^{N} p_{i,j}}{N} \right) \right\}. \tag{75}$$

- (b) In the large-system regime, the following expressions are true:

$$C_{\text{erg, can}}(\rho) - C_{\text{erg, kron}}(\rho) \rightarrow \frac{N}{2} \log_2 \left(\frac{\text{AM}_{\text{row pow}} \cdot \text{AM}_{\text{col pow}}}{\text{GM}_{\text{row pow}} \cdot \text{GM}_{\text{col pow}}} \right) \tag{76}$$

  where $\text{AM}_{\bullet}$ and $\text{GM}_{\bullet}$ correspond to the arithmetic and geometric means of row and column powers of $\mathbf{P}_c$. Further, we also have

$$0 \leq C_{\text{erg, can}}(\rho) - C_{\text{erg, kron}}(\rho) \leq 2N \log_2(N). \tag{77}$$

- (c) Equality in the lower bound holds if and only if $\mathbf{H}_c$ is regular. While it seems difficult to construct a $\mathbf{P}_c$ that meets the upper bound, the following choice is order-optimal and results in $C_{\text{erg, can}}(\rho) - C_{\text{erg, kron}}(\rho) \overset{N \to \infty}{\approx} N \log_2(N)$:

$$\mathbf{P}_c = \text{diag}\Big[N^2 - N + 1, \underbrace{1, \cdots, 1}_{N-1} \Big]. \tag{78}$$

Proof: See Appendix G. ∎

Variance of Capacity: Closed-form results are difficult to obtain for $V_\bullet(\rho)$ as $\rho \to \infty$. However, numerical studies indicate that for most scattering environments $\sqrt{V_{\text{can}}(\rho)}$ and $\sqrt{V_{\text{kron}}(\rho)}$ are sub-dominant[15] when compared with $C_{\text{erg, can}}(\rho)$ and $C_{\text{erg, kron}}(\rho)$, respectively. Thus, for a typical scattering environment, the outage capacities are primarily determined by $C_{\text{erg, can}}(\rho)$ and $C_{\text{erg, kron}}(\rho)$. The smoothing effect of the Kronecker model as can be seen from (27), the low-SNR trends of $V_\bullet(\rho)$, and numerical studies (see Figs. 2-4) lend credence to the following conjecture proving which will be the subject of future work.

Conjecture 1: The following are true for a large class of channels in the medium- to high-SNR regime:

$$\frac{\sqrt{V_{\text{can}}(\rho)}}{C_{\text{erg, can}}(\rho)} \overset{N \to \infty}{\to} 0, \ \frac{\sqrt{V_{\text{kron}}(\rho)}}{C_{\text{erg, kron}}(\rho)} \overset{N \to \infty}{\to} 0, \ \text{and} \ V_{\text{can}}(\rho) \geq V_{\text{kron}}(\rho). \tag{79}$$

Discussion: From (76), we first note that the mismatch accrued by the Kronecker model increases as the ratio of arithmetic and geometric means of the row and the column powers increases. The ratio of arithmetic and geometric means is a measure of the homogeneity of the vector (under consideration) or lack of disparities [56], [57]: The regularity of $\mathbf{P}_c$ in our context. That is, the smaller the ratio, the more regular the channel and *vice versa*. Thus, we see that the more non-regular the channel, the larger the mismatch with the Kronecker model. This conclusion is reflected in the structure of the choices of $\mathbf{P}_c$ (in Theorem 4) that lead to a large and a small mismatch. It is also reflected in Figs. 2-4 where the channels become more regular as they become richer (This is because both the transmit and the receive sides become more well-conditioned as the channel becomes richer), and the mismatch between the Kronecker and the canonical models decreases.

We also note the following trends. In the case of non-regular channels, the fact that $C_{\text{erg, can}}(\rho) > C_{\text{erg, kron}}(\rho)$ and the sub-dominance conjecture of $V_\bullet(\rho)$ implies that the Kronecker model underestimates capacity confirming the observations made in recent measurement campaigns [18], [20]–[25]. Note that the SNR range of most of these observations lie between 10 and 20 dB, which can be viewed as the high-SNR regime. The choice of the SNR range also explains why the popular belief on the decreasing probability of overestimation (see e.g., [26, Footnote 5]) has

---

[15]For example, in the i.i.d. case, it can be seen that $C_{\text{erg, can}}(\rho) = C_{\text{erg, kron}}(\rho) = \mathcal{O}(N)$ while $V_{\text{can}}(\rho) = V_{\text{kron}}(\rho) = \mathcal{O}(\log(N))$ [54].

come about. The case of regular (or near-regular) channels in the high-SNR regime has a behavior similar to that of channels in the low-SNR regime. Finally, note that the theory developed in this work is useful in the context of the probabilistic sparse framework of Sec. IV-B where it holds with probability 1. Since the class of sparse channels forms the most predominant class in the space of all possible channels (Fig. 1), the utility of this theory is immense.

VII. CONCLUSION

In this paper, we have unified existing statistical models for spatially correlated multi-antenna channels by considering a canonical decomposition of the channel along the transmit and/or the receive eigen-bases. This framework generalizes the Kronecker model, the virtual representation and the Weichselberger model, and as a by-product develops two other classes of statistical models. In addition, we have developed an abstract framework to model spatial sparsity that has been observed in many recent measurement campaigns.

These campaigns have also demonstrated that the Kronecker model results in misleading estimates for the capacity of realistic scattering environments. However, the reasons for these observations have not been well-understood so far. In this work, we have rigorously established the connection between spatial sparsity of the true channel, the non-regularity of the sparsity structure, and the impact they have on the capacity estimates provided by a Kronecker model fit. The Kronecker model fit uses the marginal sum statistics and this spreads the sparse DoF in the spatial domain. The consequent redistribution of the channel power is responsible for the mismatch in capacity estimation. In particular, we have shown that in the case of non-regular channels, the Kronecker model underestimates capacity in the medium- to high-SNR regime. On the other hand, in the low-SNR regime and regular channels in the high-SNR regime, the Kronecker model overestimates capacity at high levels of operational reliability and *vice versa*.

Possible extensions to this work include the development of a more systematic framework for the generation of correlated/sparse multi-antenna channels, the impact sparsity has on the over/underestimation of capacity and reliability, establishing rigorously the approximation in Lemma 5 and Conjecture 1, computation of closed-form expressions for the mean and the variance of capacity under the canonical and the Kronecker models at general SNRs, understanding the impact on capacity of different channel power normalizations that are consistent with physical intuition etc.

APPENDIX

A. *Proof of Lemma 4*

Consider the matrix $\mathbf{H}_c \triangleq \mathbf{U}_r^H \mathbf{H} \mathbf{U}_t$. From Assumptions 1 and 2, we can write $\mathbf{H}_c = \mathbf{U}_r^H \mathbf{H}_t$. Then, the cross-covariance of the columns of $\mathbf{H}_c$ (denoted by $\{\mathbf{h}_{ci}\}$) satisfies

$$E\left[\mathbf{h}_{ci}\,\mathbf{h}_{cj}^H\right] = \mathbf{U}_r^H E\left[\mathbf{h}_{ti}\,\mathbf{h}_{tj}^H\right]\mathbf{U}_r = \mathbf{0} \text{ for all } i, j, \ i \neq j, \tag{80}$$

which follows from the column uncorrelatedness of $\mathbf{H}_t$ in Lemma 2. Similarly, from Assumptions 3 and 4, we can write $\mathbf{H}_c = \mathbf{H}_r \mathbf{U}_t$. Then, the cross-covariance of the rows of $\mathbf{H}_c$ (denoted by $\{\mathbf{g}_{ci}\}$) satisfies

$$E\left[\mathbf{g}_{ci}\,\mathbf{g}_{cj}^H\right] = \mathbf{U}_t^H E\left[\mathbf{g}_{ri}\,\mathbf{g}_{rj}^H\right]\mathbf{U}_t = \mathbf{0} \text{ for all } i, j, \ i \neq j, \tag{81}$$

which follows from the row uncorrelatedness of $\mathbf{H}_r$ in Lemma 3. Thus, the columns and the rows of $\mathbf{H}_c$ are uncorrelated. This necessarily implies that all entries of $\mathbf{H}_c$ are uncorrelated. ∎

B. Proof of Theorem 1

Preliminaries: The following result concerning the tail probabilities of weighted sums of i.i.d. random variables would lead us towards the estimation of $I(\rho)$.

Lemma 6 (Lanzinger and Stadtmueller, [58]): Consider i.i.d. random variables $X, X_1, X_2, \cdots$ with $E[X] = 0, E[X^2] = \sigma_0^2$. Let $\beta > 0$ and $\nu \geq 2$ with $E[X^\nu] < \infty$. Define the weighted sum

$$T_n \triangleq \sum_{k=1}^{n} t_k X_k, \quad t_k \geq 0, \quad \text{and} \quad \sigma_n^2 = \sigma_0^2 \sum_{k=1}^{n} t_k^2. \tag{82}$$

Also, suppose for some $\alpha \geq 1$,

$$\frac{\max_{1 \leq k \leq n} t_k}{\sigma_n} \leq \frac{\alpha}{\sqrt{n}} \text{ for all } n. \tag{83}$$

Then, $\displaystyle \lim_{\epsilon \to 0+} \epsilon^{\nu + \frac{\nu/2-1}{\beta}} \sum_{n=1}^{\infty} n^{\nu(\beta+1/2)-2} \Pr\left(|T_n| > \epsilon n^\beta \sigma_n\right) = \frac{E\left[|\mathcal{N}|^{\nu + \frac{\nu/2-1}{\beta}}\right]}{\nu(\beta+1/2)-1} \tag{84}$

where $\mathcal{N}$ is a standard Gaussian random variable. ∎

Note that the conclusion of Lemma 6 can be suitably modified in the case of $\epsilon \nrightarrow 0+$, but is sufficiently small, by increasing the right-hand side of (84) appropriately. The crucial point is that this would not alter our conclusion since the above modification can be done, by keeping the right-hand side in (84) still finite.

Application of Lemma 6: Lemma 6 is applied in our setting as follows. Let $X_i = |\mathbf{H}_{\text{iid}}[i, j_{\max}]|^2 - 1, t_i = \mathbf{P}_c[i, j_{\max}], T_n = \sum_{i=1}^{n} \left(|\mathbf{H}_c[i, j_{\max}]|^2 - \mathbf{P}_c[i, j_{\max}]\right), \beta = \frac{1}{2}, \nu = 2$ and $n = N_r$. Then Lemma 6 implies that

$$\sum_n \Pr\left(\frac{|T_n|}{\sum_{k=1}^{n} t_k} > \eta \sqrt{\frac{\sum_{i=1}^{n} \left(\mathbf{P}_c[i, j_{\max}]\right)^2}{n}} \cdot \frac{n}{\sum_{i=1}^{n} \mathbf{P}_c[i, j_{\max}]}\right) \leq \frac{1}{\eta^2} < \infty \tag{85}$$

for η appropriately small. The conclusion in (85) implies that there exists $m > 1/2$ and $\ell > 0$ such that

$$\Pr\left(\left|\sum_{i=1}^{N_r} |\mathbf{H}_c[i, j_{\max}]|^2 - \mathbf{P}_c[i, j_{\max}]\right| > \eta \sqrt{N_r \sum_{i=1}^{N_r} \left(\mathbf{P}_c[i, j_{\max}]\right)^2}\right) \leq \frac{1}{\eta^{2\ell} N_r^{2m}}. \tag{86}$$

Proof of Theorem: Let $\mathbf{Y} = \mathbf{Y}(N_r)$ denote the random variable $\sum_{i=1}^{N_r} \left| \mathbf{H}_c[i, j_{\max}] \right|^2$. Setting $\rho = \frac{1}{\sum_{i=1}^{N_r} \mathbf{P}_c[i,j_{\max}]} \cdot \frac{\eta}{1-\eta} \cdot \frac{1}{\gamma}$ with $\gamma = 1 + \eta \cdot \frac{\sqrt{N_r \sum_{i=1}^{N_r} (\mathbf{P}_c[i,j_{\max}])^2}}{\sum_{i=1}^{N_r} \mathbf{P}_c[i,j_{\max}]}$ and using (86), we get

$$\Pr\left(\rho\mathbf{Y} > \frac{\eta}{1-\eta} \right) = \Pr\left(\rho\mathbf{Y} - \rho\sum_{i=1}^{N_r}\mathbf{P}_c[i, j_{\max}] > \frac{\eta}{1-\eta} - \rho\sum_{i=1}^{N_r}\mathbf{P}_c[i, j_{\max}] \right) \quad (87)$$

$$\leq \Pr\left(\rho\left|\mathbf{Y} - E[\mathbf{Y}]\right| > \eta\rho\sqrt{N_r\sum_{i=1}^{N_r}\left(\mathbf{P}_c[i, j_{\max}]\right)^2} \right) \leq \frac{1}{\eta^{2\ell} N_r^{2m}}.$$

An upper bound to $I(\rho)$ follows easily from the log-inequality:

$$I(\rho) \triangleq E\left[\log_2\left(1 + \rho\mathbf{Y}\right)\right] \leq \rho\log_2(e)\,E\left[\mathbf{Y}\right] = \rho\log_2(e)\sum_i \mathbf{P}_c[i, j_{\max}]. \quad (88)$$

We now establish a tight lower bound for $I(\rho)$:

$$\frac{E\left[\log_2\left(1 + \rho\mathbf{Y}\right)\right]}{\log_2(e)} \overset{(a)}{\geq} E\left[\frac{\rho\mathbf{Y}}{1 + \rho\mathbf{Y}}\right] \quad (89)$$

$$= \underbrace{E\left[\frac{\rho\mathbf{Y}}{1 + \rho\mathbf{Y}}\,\chi\left(\rho\mathbf{Y} \leq \frac{\eta}{1-\eta}\right)\right]}_{Z_1} + \underbrace{E\left[\frac{\rho\mathbf{Y}}{1 + \rho\mathbf{Y}}\,\chi\left(\rho\mathbf{Y} > \frac{\eta}{1-\eta}\right)\right]}_{Z_2} \quad (90)$$

$$\overset{(b)}{\geq} \underbrace{E\left[\left(\rho\mathbf{Y} - \rho^2\mathbf{Y}^2\right)\chi\left(\rho\mathbf{Y} \leq \frac{\eta}{1-\eta}\right)\right]}_{Z_3} \quad (91)$$

$$Z_3 = E\left[\rho\mathbf{Y}\right] - \underbrace{E\left[\rho\mathbf{Y}\,\chi\left(\rho\mathbf{Y} > \frac{\eta}{1-\eta}\right)\right]}_{Z_4} - \underbrace{E\left[\rho^2\mathbf{Y}^2\,\chi\left(\rho\mathbf{Y} \leq \frac{\eta}{1-\eta}\right)\right]}_{Z_5}$$

where (a) follows from the inequality $\log_e(1+z) \geq \frac{z}{1+z}$ and (b) follows from using the inequality $\frac{1}{1+z} \geq (1-z)$. An application of the Cauchy-Schwarz inequality shows that

$$\frac{E\left[\log_2\left(1 + \rho\mathbf{Y}\right)\right]}{\log_2(e)} \geq \rho \cdot \sum_{i=1}^{N_r}\mathbf{P}_c[i, j_{\max}]\underbrace{\left[1 - \frac{\sqrt{E[\mathbf{Y}^2]}}{E[\mathbf{Y}]} \cdot \frac{1}{\eta^\ell N_r^m} - \rho \cdot \frac{E[\mathbf{Y}^2]}{E[\mathbf{Y}]}\right]}_{Z_6}. \quad (92)$$

Now, the quantity Z_6 makes meaningful sense as a lower bound to $I(\rho)$ only if it is positive. Plugging in the expression for ρ, we see that this can be ensured[16] *if η is constrained to $(0, 1/2]$.* Further, $\eta = 1/2$ maximizes the lower bound to $I(\rho)$. Evaluating $E[\mathbf{Y}^2]$, substituting the value of κ_c and noting that $\rho = \frac{\delta}{\sum_{i=1}^{N_r}\mathbf{P}_c[i,j_{\max}]}$ for $\delta < \frac{1}{\gamma_0}$ where γ_0 is as in (40), we get

$$I(\rho) \geq \log_2(e) \cdot \delta \cdot \left(1 - \frac{2^\ell\sqrt{\kappa_c}}{N_r^m} - \frac{\delta\kappa_c}{\gamma_0}\right). \quad (93)$$

---

[16]The choice of $1/2$ for the upper bound of the valid interval of η is more or less arbitrary and we have not optimized over this choice.

To complete the proof, we observe from [47] that beamforming is the optimal signaling strategy for all $\rho < \rho_{\text{low, can}}$ and there exists a constant $d > 0$ (independent of $\mathbf{P}_c$, N_r and N_t) such that

$$\rho_{\text{low, can}} \geq \frac{d}{\sum_{i=1}^{N_r} \mathbf{P}_c[i, j_{\max}]}. \tag{94}$$

The constant c in the statement of the theorem can be chosen to be $\min(d, 1/\gamma_0)$. It is important to note that the tightness of the upper bound in (88) and the lower bound in (89) critically hinge on the low-SNR assumption. Thus the theorem is complete. ∎

C. Proof of Theorem 2

- (a) From (35), the ergodic capacities in the low-SNR regime are given by

$$C_{\text{erg, can}}(\rho) = E\left[\log_2\left(1 + \rho \sum_i \left|\mathbf{H}_c[i, j_{\max}]\right|^2\right)\right], \tag{95}$$

$$C_{\text{erg, kron}}(\rho) = E\left[\log_2\left(1 + \rho \sum_i \left|\mathbf{H}_k[i, j_{\max}]\right|^2\right)\right] \tag{96}$$

where $j_{\max} = \arg\max_j \sum_i \mathbf{P}_c[i, j] = \arg\max_j \sum_i \mathbf{P}_k[i, j]$. Estimating these quantities is a straightforward consequence of Theorem 1.

- (b) For the variance, we have

$$V_{\text{can}}(\rho) \triangleq E\left[(\log_2(1 + \rho \mathbf{Y}))^2\right] - (E[\log_2(1 + \rho \mathbf{Y})])^2 \tag{97}$$

where $\mathbf{Y} = \sum_{i=1}^{N_r} \left|\mathbf{H}_c[i, j_{\max}]\right|^2$. Proceeding along similar lines as in App. B, we have

$$V_{\text{can}}(\rho) \leq (\log_2(e))^2 \cdot \delta^2 \cdot \left(\kappa_c - \left(1 - \frac{2^\ell \sqrt{\kappa_c}}{N_r^m} - \frac{\delta \kappa_c}{\gamma_0}\right)^2\right) \tag{98}$$

$$V_{\text{can}}(\rho) \geq (\log_2(e))^2 \cdot \delta^2 \cdot \left(\kappa_c - 1 - \frac{\sqrt{E[\mathbf{Y}^4]} \cdot 2^\ell}{\left(\sum_{i=1}^{N_r} \mathbf{P}_c[i, j_{\max}]\right)^2 \cdot N_r^m} - \frac{2\delta \cdot E[\mathbf{Y}^3]}{\left(\sum_{i=1}^{N_r} \mathbf{P}_c[i, j_{\max}]\right)^3}\right)$$

where the constants are as in the statement of Theorem 1. Thus, we can recast $V_{\text{can}}(\rho)$ as

$$V_{\text{can}}(\rho) = (\log_2(e))^2 \cdot \rho^2 \cdot \sum_i \left(\mathbf{P}_c[i, j_{\max}]\right)^2 \cdot (1 + o(1)) \tag{99}$$

where the $o(1)$ factor in the above expression converges to 0 as $N_r \to \infty$ and $\rho \to 0$. The critical assumption in the above proof is that $\mathbf{H}_c[i, j]$ are independent random variables. Thus, the same proof technique can be adapted to compute $V_{\text{kron}}(\rho)$ as well.

- (c) The relationship between $V_{\text{can}}(\rho)$ and $V_{\text{kron}}(\rho)$ is not obvious. For this, we need the following result on the monotonicity of ratios of means [55, pp. 129-130].

Lemma 7 (Marshall, Olkin and Proschan): Let $\mathbf{x} = [x_1, \cdots, x_n]$ and $\mathbf{y} = [y_1, \cdots, y_n]$ be two vectors such that $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$. If $\frac{y_i}{x_i}$ is decreasing in i and $x_1 \geq \cdots \geq x_n > 0$, then $\mathbf{x}$ is majorized by $\mathbf{y}$, and

$$g(r) \triangleq \left(\frac{\sum_{i=1}^{n} x_i^r}{\sum_{i=1}^{n} y_i^r} \right)^{1/r} \tag{100}$$

is decreasing in r for $r > 0$. ∎

Application of Lemma 7: We set

$$x_i \triangleq \frac{\sum_{k=1}^{N_t} \mathbf{P}_c[i, k]}{\rho_c}, \qquad y_i \triangleq \frac{\mathbf{P}_c[i, j_{\max}]}{\sum_{k=1}^{N_r} \mathbf{P}_c[k, j_{\max}]} \tag{101}$$

and $n = N_r$. From the assumption in the statement of the theorem, note that $x_i > 0$ for all i and are in decreasing order. The fact that $\frac{y_i}{x_i}$ is decreasing is a consequence of (48). A straightforward consequence of Lemma 7 is that $V_{\text{can}}(\rho) \geq V_{\text{kron}}(\rho)$. ∎

D. Proof of Prop. 2

- (a) With framework II, the main goal is to compute the probability of failure of (48). Towards this computation, we first condition upon $p_{i,j_{\max}}$ and $p_{i+1,j_{\max}}$ (in particular, $s_\bullet$ and $q_\bullet$) where i is such that $1 \leq i \leq N_r - 1$. Define the conditional probability p_i:

$$p_i \triangleq \text{Pr}\left(\frac{\sum_{k \neq j_{\max}} \mathbf{P}_c[i, k]}{\sum_{k \neq j_{\max}} \mathbf{P}_c[i+1, k]} > \frac{\mathbf{P}_c[i, j_{\max}]}{\mathbf{P}_c[i+1, j_{\max}]} \right) \tag{102}$$

$$= \text{Pr}\left(p_{i+1,j_{\max}} \sum_{k \neq j_{\max}} q_{i,k} s_{i,k} - p_{i,j_{\max}} \sum_{k \neq j_{\max}} q_{i+1,k} s_{i+1,k} > 0 \right). \tag{103}$$

Hence, the conditional probability of failure of (48) is $1 - \prod_{i=1}^{N_r - 1}(1 - p_i)$. We intend to show that the above probability converges to 0 as $N_t \to \infty$.

Without loss in generality, we can assume that $s_{i,j_{\max}} = s_{i+1,j_{\max}} = 1$ (Otherwise, $p_i = 0$.). Similarly, we can assume that $\{q_{i,j_{\max}}\}$ is decreasing in i. Note that p_i can be written as

$$p_i = \text{Pr}\left(\frac{\mathbf{Z}_{N_t}}{N_t} > 0 \right) \quad \text{where} \tag{104}$$

$$\mathbf{Z}_{N_t} = q_{i+1,j_{\max}} \sum_{k \neq j_{\max}} q_{i,k} s_{i,k} - q_{i,j_{\max}} \sum_{k \neq j_{\max}} q_{i+1,k} s_{i+1,k}. \tag{105}$$

Using the independence of $\{q_{i,k}\}$ and $\{s_{i,k}\}$ and their statistics, it can be checked that

$$\frac{E[\mathbf{Z}_{N_t}]}{N_t} \overset{N_t \to \infty}{\longrightarrow} (q_{i+1,j_{\max}} - q_{i,j_{\max}}) p\mu < 0, \tag{106}$$

$$\text{Var}\left(\frac{\mathbf{Z}_{N_t}}{N_t} \right) \overset{N_t \to \infty}{\longrightarrow} \frac{\left((q_{i+1,j_{\max}})^2 + (q_{i,j_{\max}})^2 \right) p\sigma^2}{N_t} \to 0. \tag{107}$$

That is, $\frac{\mathbf{Z}_{N_t}}{N_t}$ hardens around its mean (which is negative) as $N_t \to \infty$ and hence, p_i converges to 0. Averaging over $\{p_{i,j_{\max}}\}$, we see that for "almost all" sparse scattering environments, the condition in (48) holds and hence, $V_{\mathsf{can}}(\rho) \geq V_{\mathsf{kron}}(\rho)$.

- (b) We now compare the dominant terms of the variances of capacity with the two models. We have

$$\frac{V_{\mathsf{can}}(\rho)}{N_r \ (\log_2(e)\rho)^2} = \frac{\sum_i p_{i,j_{\max}}^2}{N_r} \cdot \frac{1}{\left(\frac{\sum_{kl} p_{k,l}}{N_t N_r}\right)^2}. \tag{108}$$

After using (27), we can also write $V_{\mathsf{kron}}(\rho)$ in terms of $\{p_{i,j}\}$ as

$$\frac{V_{\mathsf{kron}}(\rho)}{N_r \ (\log_2(e)\rho)^2} = \frac{\left(\frac{\sum_k p_{k,j_{\max}}}{N_r}\right)^2 \cdot \frac{1}{N_r} \sum_i \left(\frac{\sum_l p_{i,l}}{N_t}\right)^2}{\left(\frac{\sum_{kl} p_{k,l}}{N_t N_r}\right)^4}. \tag{109}$$

In the large-system regime, since $\{p_{i,j} = q_{i,j} s_{i,j}\}$ and $\{q_{i,j}\}$ is a realization from an i.i.d. family of mean μ and variance σ^2, we can use the law of large numbers [59] to check that

$$\frac{V_{\mathsf{can}}(\rho)}{N_r \ (\log_2(e)\rho)^2} \to \frac{E[(q_{i,j})^2]}{(E[q_{i,j}])^2 \, p}, \qquad \frac{V_{\mathsf{kron}}(\rho)}{N_r \ (\log_2(e)\rho)^2} \to 1. \tag{110}$$

The fact that $V_{\mathsf{can}}(\rho) \geq V_{\mathsf{kron}}(\rho)$ follows from the Cauchy-Schwarz inequality. The upper bound for $\frac{V_{\mathsf{can}}(\rho)}{V_{\mathsf{kron}}(\rho)}$ follows from the reverse Cauchy-Schwarz inequality [60, equation 24, p. 208] due to Cassels, which is stated here for convenience.

Lemma 8: If $\mathbf{x} = [x_1, \cdots, x_n]$, $\mathbf{y} = [y_1, \cdots, y_n]$ and $\mathbf{w} = [w_1, \cdots, w_n]$ are positive n-tuples such that $0 < m_1 \leq x_i \leq M_1$ and $0 < m_2 \leq y_i \leq M_2$ for all i with $m_1 m_2 < M_1 M_2$, then

$$\frac{\left(\sum_{i=1}^n x_i^2 w_i^2\right) \cdot \left(\sum_{i=1}^n y_i^2 w_i^2\right)}{\left(\sum_{i=1}^n x_i y_i w_i^2\right)^2} \leq \frac{(m_1 m_2 + M_1 M_2)^2}{4 m_1 m_2 M_1 M_2}. \tag{111}$$

∎

- (c) Equality in the lower bound is possible if and only if $q_{i,j}$ is constant with probability 1 and $p = 1$. That is, $\mathbf{H}_c$ and $\mathbf{H}_k$ are i.i.d. With $\mathbf{P}_c$ as in (52), it can be checked that

$$\frac{V_{\mathsf{can}}(\rho)}{V_{\mathsf{kron}}(\rho)} = \frac{(M+m)^2}{4Mm} \cdot \frac{1 + \frac{2m+M}{mN}}{\left(1 + \frac{M+m}{2mN}\right)^2} \stackrel{N \to \infty}{\to} \frac{(M+m)^2}{4Mm}. \tag{112}$$

Thus the proposition is complete. ∎

E. Proof of Theorem 3

When $r > 1$, we assume that the r dominant columns have been relabeled as columns 1 through r. We then have

$$\frac{E_b}{N_{o\,\min,\,\mathsf{can}}} = \frac{E_b}{N_{o\,\min,\,\mathsf{kron}}} = \frac{r \log_e(2)}{\sum_{i=1}^{N_r} \sum_{j=1}^r \mathbf{P}_c[i,j]} = \frac{\log_e(2)}{\sum_{i=1}^{N_r} \mathbf{P}_c[i, j_{\max}]} \tag{113}$$

with the last equality following because all the r columns have the same sums. Using the uniform input over r modes and the Gaussian moment factoring theorem, the wideband slopes can be checked to be

$$S_{0,\,\mathsf{can}} \;=\; 2 \cdot \frac{\left(\sum_{i=1}^{N_r}\sum_{j=1}^{r} \mathbf{P}_c[i,j]\right)^2}{\sum_{i=1}^{N_r}\sum_{j_1=1}^{r}\sum_{j_2=1}^{r}\mathbf{P}_c[i,j_1]\mathbf{P}_c[i,j_2] + \sum_{j=1}^{r}\left(\sum_{i=1}^{N_r}\mathbf{P}_c[i,j]\right)^2}, \qquad (114)$$

$$S_{0,\,\mathsf{kron}} \;=\; 2 \cdot \frac{\left(\sum_{i=1}^{N_r}\sum_{j=1}^{r} \mathbf{P}_k[i,j]\right)^2}{\sum_{i=1}^{N_r}\sum_{j_1=1}^{r}\sum_{j_2=1}^{r}\mathbf{P}_k[i,j_1]\mathbf{P}_k[i,j_2] + \sum_{j=1}^{r}\left(\sum_{i=1}^{N_r}\mathbf{P}_k[i,j]\right)^2}. \qquad (115)$$

Using the law of large numbers appropriately, we have

$$S_{0,\,\mathsf{can}} \to \frac{2N_r r \mu^2 p}{\mu^2(N_r p + (r-1)p + 1) + \sigma^2}, \quad \text{and} \quad S_{0,\,\mathsf{kron}} \to \frac{2N_r r}{N_r + r}. \qquad (116)$$

$\blacksquare$

F. Proof of Lemma 5

See [44, Chap. 2, p. 104] for a proof of the first statement. For the statement on determinant approximation, we start with [61, p. 35, 39] which states that $\det\left(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H\right)$ can be decomposed as a product of independent random variables, $\widehat{\mathbf{Z}}_i$ where

$$\widehat{\mathbf{Z}}_i \sim \sum_{j=1}^{i} |\eta_{ij}|^2, \quad \eta_{ij} = \sum_{l=1}^{N} \widetilde{\mathbf{H}}[i,l]\,\theta[l,j], \qquad (117)$$

and the matrix $\Theta = \{\theta[i,j]\}$ is a unitary random matrix independent of $\widetilde{\mathbf{H}}$. Note that

$$|\eta_{ij}|^2 = \sum_{l}\left|\widetilde{\mathbf{H}}[i,l]\right|^2 |\theta[l,j]|^2 + \sum_{l_1 \neq l_2} \widetilde{\mathbf{H}}[i,l_1]\widetilde{\mathbf{H}}[i,l_2]^\star \theta[l_1,j]\theta[l_2,j]^\star \qquad (118)$$

and using the facts that the entries of a random unitary matrix are asymptotically self-averaging, (that is, zero mean in a "statistical" sense) and the rows and columns have unit norm, we have the following approximation for $|\eta_{ij}|^2$:

$$|\eta_{ij}|^2 \approx \frac{1}{N}\sum_{l}\left|\widetilde{\mathbf{H}}[i,l]\right|^2. \qquad (119)$$

This leads to the approximation for $\widehat{\mathbf{Z}}_i$, which we denote by $\widetilde{\mathbf{Z}}_i$ in the statement of lemma. $\blacksquare$

G. *Proof of Theorem 4*

- (a) The i.i.d. result can be exploited in the Kronecker case as follows:

$$\log_2 \det \left(\mathbf{H}_k \, \mathbf{H}_k^H \right) \;\overset{(a)}{=}\; \log_2 \det \left(\mathbf{\Lambda}_r \, \mathbf{H}_{\mathsf{iid}} \, \mathbf{\Lambda}_t \, \mathbf{H}_{\mathsf{iid}}^H \right) \tag{120}$$

$$\overset{(b)}{=}\; \log_2 \det(\mathbf{\Lambda}_t \, \mathbf{\Lambda}_r) + \log_2 \det \left(\mathbf{H}_{\mathsf{iid}} \, \mathbf{H}_{\mathsf{iid}}^H \right) \tag{121}$$

$$\overset{(c)}{\sim}\; \sum_{i=1}^{N} \log_2 \left(\frac{1}{2} \cdot \mathbf{P}_k[i] \cdot \chi^2 \left(2(N - i + 1) \right) \right) \tag{122}$$

where (a) follows from the definition of $\mathbf{H}_k$, (b) from the fact that $N_t = N_r$ and $\det(\mathbf{AB}) = \det(\mathbf{BA})$, and (c) from (67) and the definitions of $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_r$. Using $E\left[\log_2 \det(\mathbf{H}_{\mathsf{iid}}\mathbf{H}_{\mathsf{iid}}^H) \right]$ [54], we can compute $C_{\mathsf{erg, \, kron}}(\rho)$ to be

$$C_{\mathsf{erg, \, kron}}(\rho) \;\rightarrow\; N \log_2 \left(\frac{\rho \, N^2}{\sum_{ij} p_{i,j}} \right) + \sum_{i=1}^{N} \log_2 \left(\frac{i}{N} \right) + K_{\mathsf{kron}} + \mathcal{O}\!\left(\frac{1}{\rho} \right) \tag{123}$$

where

$$K_{\mathsf{kron}} \;=\; \sum_{i=1}^{N} \log_2 \left(\frac{\sum_l p_{i,l} \sum_k p_{k,i}}{\sum_{kl} p_{k,l}} \right). \tag{124}$$

For the canonical case, we write $\mathbf{H}_c \mathbf{H}_c^H$ as

$$\mathbf{H}_c \mathbf{H}_c^H = \frac{N^2}{\sum_{ij} p_{i,j}} \widetilde{\mathbf{H}} \widetilde{\mathbf{H}}^H, \; \widetilde{\mathbf{H}}[i, j] \sim \mathcal{CN}(0, p_{i,j}) \tag{125}$$

and compute $C_{\mathsf{erg, \, can}}(\rho)$ as follows:

$$E\left[\log_2 \det(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H) \right] \;\overset{(a)}{\approx}\; \sum_{i=1}^{N} \log_2(i) + \sum_{i=1}^{N} E\left[\log_2 \left(\frac{\sum_{j=1}^{N} |\widetilde{\mathbf{H}}[i, j]|^2}{N} \right) \right] \tag{126}$$

$$\overset{(b)}{\rightarrow}\; \sum_{i=1}^{N} \log_2(i) + \underbrace{\sum_{i=1}^{N} \log_2 \left(\frac{\sum_{j=1}^{N} p_{i,j}}{N} \right)}_{K_{\mathsf{can}}} \tag{127}$$

$$C_{\mathsf{erg, \, can}}(\rho) \;\approx\; N \log_2 \left(\frac{\rho \, N^2}{\sum_{ij} p_{i,j}} \right) + \sum_{i=1}^{N} \log_2 \left(\frac{i}{N} \right) + K_{\mathsf{can}} + \mathcal{O}\!\left(\frac{1}{\rho} \right)$$

where (a) follows from the approximation (the approximation gets more accurate as $N \rightarrow \infty$) in Lemma 5. The convergence in (b) follows from Prop. 3 which is stated and proved next. Since $N_r = N_t = N$, all the above steps are true even if $\mathbf{H}_c$ is replaced with $\mathbf{H}_c^H$. This leads to the expression for K_{can} in (75).

Proposition 3: With the setting as above, we have

$$\log_2 \left(\frac{\sum_{j=1}^{N} |\widetilde{\mathbf{H}}[i, j]|^2}{N} \right) \;\overset{N \rightarrow \infty}{\rightarrow}\; \log_2 \left(\frac{\sum_{j=1}^{N} p_{i,j}}{N} \right) \text{ in mean for any } i. \tag{128}$$

Proof: We decompose the left-hand side as

$$E\left[\log_2\left(\frac{\sum_{j=1}^{N}|\widetilde{\mathbf{H}}[i,j]|^2}{N}\right)\right]$$

$$= E\left[\log\left(\frac{\sum_{j=1}^{N}|\widetilde{\mathbf{H}}[i,j]|^2\chi\left(|\widetilde{\mathbf{H}}[i,j]|^2\leq K\right)+|\widetilde{\mathbf{H}}[i,j]|^2\chi\left(|\widetilde{\mathbf{H}}[i,j]|^2> K\right)}{N}\right)\right]$$

$$= E\left[\log\left(\frac{\sum_{j=1}^{N}|\widetilde{\mathbf{H}}[i,j]|^2\chi\left(|\widetilde{\mathbf{H}}[i,j]|^2\leq K\right)}{N}\right)\right]$$

$$+ E\left[\log\left(1+\frac{\sum_{j=1}^{N}|\widetilde{\mathbf{H}}[i,j]|^2\chi\left(|\widetilde{\mathbf{H}}[i,j]|^2> K\right)}{\sum_{j=1}^{N}|\widetilde{\mathbf{H}}[i,j]|^2\chi\left(|\widetilde{\mathbf{H}}[i,j]|^2\leq K\right)}\right)\right] \qquad (129)$$

for some $K>0$ fixed.

For the first term, note that the weak law of large numbers states that for all i

$$\frac{\sum_{j=1}^{N}|\widetilde{\mathbf{H}}[i,j]|^2\chi\left(|\widetilde{\mathbf{H}}[i,j]|^2\leq K\right)}{N} \xrightarrow[N\to\infty]{} \frac{\sum_{j=1}^{N}E\left[|\widetilde{\mathbf{H}}[i,j]|^2\chi\left(|\widetilde{\mathbf{H}}[i,j]|^2\leq K\right)\right]}{N} \qquad (130)$$

$$= P,$$

$$P \triangleq \frac{\sum_{j=1}^{N}p_{i,j}-(p_{i,j}+K)\,e^{-\frac{K}{p_{i,j}}}}{N}. \qquad (131)$$

The convergence is in probability and hence, also weakly [62, p. 310]. The second equality follows from a routine expectation computation. Since $\log(\cdot)$ is a continuous function and the limit random variable is a constant, following [62, p. 316, p. 310] we also have

$$\log\left(\frac{\sum_{j=1}^{N}|\widetilde{\mathbf{H}}[i,j]|^2\chi\left(|\widetilde{\mathbf{H}}[i,j]|^2\leq K\right)}{N}\right) \xrightarrow{p} \log\left(P\right). \qquad (132)$$

The above convergence can further be strengthened to convergence in mean since the random variables are bounded by K for all i and all choices of N [62, p. 310].

For the second term, we use the following lower bound:

$$|\widetilde{\mathbf{H}}[i,j]|^2\chi\left(|\widetilde{\mathbf{H}}[i,j]|^2\leq K\right) \geq |\widetilde{\mathbf{H}}[i,j]|^2\chi\left(\epsilon< |\widetilde{\mathbf{H}}[i,j]|^2\leq K\right) \qquad (133)$$

for some $0 < \epsilon \leq K$. Using this, we can upper bound the second term by

$$E\left[\log\left(1 + \frac{\sum_{j=1}^{N} |\widetilde{\mathbf{H}}[i,j]|^2 \chi\left(|\widetilde{\mathbf{H}}[i,j]|^2 > K\right)}{\sum_{j=1}^{N} |\widetilde{\mathbf{H}}[i,j]|^2 \chi\left(\epsilon < |\widetilde{\mathbf{H}}[i,j]|^2 \leq K\right)}\right)\right]$$

$$\leq E\left[\log\left(1 + \frac{\sum_{j=1}^{N} |\widetilde{\mathbf{H}}[i,j]|^2 \chi\left(|\widetilde{\mathbf{H}}[i,j]|^2 > K\right)}{N\epsilon}\right)\right] \tag{134}$$

$$\leq \frac{\sum_{j=1}^{N} E\left[|\widetilde{\mathbf{H}}[i,j]|^2 \chi\left(|\widetilde{\mathbf{H}}[i,j]|^2 > K\right)\right]}{N\epsilon} \tag{135}$$

$$= \frac{\sum_{j=1}^{N} (p_{i,j} + K) e^{-\frac{K}{p_{i,j}}}}{N\epsilon} \tag{136}$$

where the second step follows from the log-inequality. Combining these two results by choosing K sufficiently large to ensure that $(p_{i,j} + K) e^{-\frac{K}{p_{i,j}}}$ is sufficiently small for all i, j and ϵ finite, we obtain the conclusion as in the statement of the proposition. ∎

- (b) In the large-system regime, we have

$$C_{\text{erg, can}}(\rho) - C_{\text{erg, kron}}(\rho) \approx K_{\text{can}} - K_{\text{kron}} = \sum_{i=1}^{N} \log_2\left(\frac{\sum_{kl} p_{k,l}}{N\sqrt{\sum_l p_{i,l} \sum_k p_{k,i}}}\right) \tag{137}$$

$$= \log_2\left(\frac{N^N}{\left(\prod_{i=1}^{N} P_i Q_i\right)^{1/2}}\right) \tag{138}$$

$$= \frac{N}{2} \log_2\left(\frac{\text{AM}_{\text{row pow}} \cdot \text{AM}_{\text{col pow}}}{\text{GM}_{\text{row pow}} \cdot \text{GM}_{\text{col pow}}}\right) \tag{139}$$

where $P_i = \sum_{j=1}^{N} \mathbf{P}_c[j,i]$ and $Q_i = \sum_{j=1}^{N} \mathbf{P}_c[i,j]$ are the column and the row powers, respectively such that $\sum_i P_i = \sum_i Q_i = N^2$.

An application of the arithmetic-geometric mean inequality shows that $K_{\text{can}} \geq K_{\text{kron}}$. For an upper bound on the difference, we use the reverse arithmetic-geometric mean inequality [60, Theorem 3, p. 124] due to Docev, which is stated here for convenience.

Lemma 9: If $\mathbf{x} = [x_1, \cdots, x_n]$ is a positive n-tuple with $K = \frac{\max_i x_i}{\min_i x_i}$, then

$$\frac{\text{AM}_{\mathbf{x}}}{\text{GM}_{\mathbf{x}}} \leq \frac{(K-1)K^{\frac{1}{K-1}}}{e \log(K)}. \tag{140}$$

∎

Since $\mathbf{P}_c$ is rank-N, we apply Lemma 9 with $K = N^2 - N + 1$ for an upper bound, and the result is (77).

- (c) Equality in the application of the arithmetic-geometric mean inequality is possible if and only if $P_i = Q_i = N$ for all i. It is straightforward to check that a channel satisfying

this property has to be necessarily regular (see Footnote 3). The conclusion for the lower bound follows by plugging the choice of $\mathbf{P}_c$ in (78) in the capacity expressions.

∎

REFERENCES

[1] Í. E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–596, Nov. 1999.

[2] G. J. Foschini, "Layered Space-Time Architechture for Wireless Communication in a Fading Environment when Using Multi-Element Antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[3] C-N. Chuah, J. M. Kahn, and D. N. C. Tse, "Capacity Scaling in MIMO Wireless Systems under Correlated Fading," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[4] D-S. Shiu, G. J. Foschini, M. Gans, and J. M. Kahn, "Fading Correlation and Its Effect on the Capacity of Multielement Antenna Systems," *IEEE Trans Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[5] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO Wireless Channels: Models and Performance Prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[6] A. M. Sayeed, "Deconstructing Multi-Antenna Fading Channels," *IEEE Trans. Sig. Proc.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[7] K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity Scaling and Spectral Efficiency in Wideband Correlated MIMO Channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2504–2526, Oct. 2003.

[8] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated MIMO Rayleigh Fading Channels: Capacity, Optimal Signaling and Asymptotics," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2058–2072, June 2005.

[9] A. S. Y. Poon, R. W. Broderson, and D. N. C. Tse, "Degrees of Freedom in Multiple-Antenna Channels: A Signal Space Approach," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 523–536, Feb. 2005.

[10] K. Yu, M. Bengtsson, B. Ottersten, P. Karlsson, and M. A. Beach, "Second Order Statistics of NLOS Indoor MIMO Channels Based on 5.2 GHz Measurements," *IEEE Global Telecommun. Conf.*, vol. 1, pp. 156–160, Nov. 2001.

[11] D. P. McNamara, M. A. Beach, and P. N. Fletcher, "Spatial Correlation in Indoor MIMO Channels," *IEEE Intern. Symp. Pers. Ind. Mob. Radio Commun.*, vol. 1, pp. 290–294, Sept. 2002.

[12] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A Stochastic MIMO Radio Channel Model with Experimental Validation," *IEEE Journ. Sel. Areas in Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[13] J. W. Wallace, M. A. Jensen, A. L. Swindlehurst, and B. D. Jeffs, "Experimental Characterization of the MIMO Wireless Channel: Data Acquisition and Analysis," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 335–343, Mar. 2003.

[14] D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-Input-Multiple-Output Measurements and Modeling in Manhattan," *IEEE Journ. Sel. Areas in Commun.*, vol. 21, no. 3, pp. 321–331, Apr. 2003.

[15] J. H. Kotecha and A. M. Sayeed, "Optimal Signal Design for Estimation of Correlated MIMO Channels," *IEEE Intern. Conf. Commun.*, vol. 5, pp. 3170–3174, May 2003.

[16] J. H. Kotecha and A. M. Sayeed, "Transmit Signal Design for Optimal Estimation of Correlated MIMO Channels," *IEEE Trans. Sig. Proc.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[17] J. H. Kotecha and A. M. Sayeed, "Canonical Statistical Models for Correlated MIMO Fading Channels and Capacity Analysis," *Technical Report #ECE-03-05, University of Wisconsin-Madison*, Mar. 2004, Available: [Online]. http://dune.ece.wisc.edu/.

[18] W. Weichselberger, M. Herdin, H. Özcelik, and E. Bonek, "A Stochastic MIMO Channel Model with Joint Correlation of Both Link Ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, Jan. 2006.

[19] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of Antenna Correlation on the Capacity of Multiantenna Channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.

[20] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental Study of MIMO Channel Statistics and Capacity via the Virtual Channel Representation," *Technical Report, University of Wisconsin-Madison*, Feb. 2007, Available: [Online]. `http://dune.ece.wisc.edu/`.

[21] N. Costa and S. Haykin, "A Novel Wideband Channel Model and Experimental Validation," *IEEE Trans. Antennas and Propagat.*, vol. 56, no. 2, pp. 550–562, Feb. 2008.

[22] H. Ozcelik, N. Czink, and E. Bonek, "What Makes a Good MIMO Channel Model?," *IEEE Spring Veh. Tech. Conf.*, vol. 1, pp. 156–160, May 2005.

[23] S. Wyne, A. Molisch, P. Almers, G. Eriksson, J. Karedal, and F. Tufvesson, "Statistical Evaluation of Outdoor-to-Indoor Office MIMO Measurements at 5.2 GHz," *IEEE Fall Veh. Tech. Conf.*, vol. 1, pp. 146–150, May 2005.

[24] T. A. Lamahewa, R. A. Kennedy, T. D. Abhayapala, and T. Betlehem, "MIMO Channel Correlation in General Scattering Environments," *Aus. Commun. Theory Workshop*, pp. 93–98, Feb. 2006.

[25] P. Almers, E. Bonek, and A. Burr et al., "Survey of Channel and Radio Propagation Models for Wireless MIMO Systems," *EURASIP Journ. Wireless Commun. and Networking*, vol. 2007, 2007.

[26] E. Bonek, "Experimental Validation of Analytical MIMO Channel Models," *Elektrotechnik und Informationstechnik (e&i)*, vol. 122, no. 6, pp. 196–205, 2005.

[27] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-Achieving Input Covariance for Correlated Multi-Antenna Channels," *Allerton Conf. Commun. Cont. and Comp.*, 2003.

[28] L. Wood and W. S. Hodgkiss, "A Reduced-Rank Eigenbasis MIMO Channel Model," *IEEE Wireless Telecommun. Symp.*, pp. 78–83, Apr. 2008.

[29] O. M. Bucci and G. Franceschetti, "On Degrees of Freedom of Scattered Fields," *IEEE Trans. Antennas and Propagat.*, vol. 37, no. 7, pp. 918–926, July 1989.

[30] M. D. Migliore, "On the Role of the Number of Degrees of Freedom of the Field in MIMO Channels," *IEEE Trans. Antennas and Propagat.*, vol. 54, no. 2, pp. 620–628, Feb. 2006.

[31] K. Chakraborty and M. Franceschetti, "Maxwell Meets Shannon: Space-Time Duality in Multiple Antenna Channels," *Allerton Conf. Commun. Cont. and Comp.*, 2006.

[32] J. Xu and R. Janaswamy, "Electromagnetic Degrees of Freedom in 2-D Scattering Environments," *IEEE Trans. Antennas and Propagat.*, vol. 54, no. 12, pp. 3882–3894, Dec. 2006.

[33] L. W. Hanlen, R. Timo, and R. Perera, "On Dimensionality for Sparse Multipath," *Aus. Commun. Theory Workshop*, pp. 125–129, Feb. 2006.

[34] H. M. Jones, R. A. Kennedy, and T. D. Abhayapala, "On Dimensionality of Multipath Fields: Spatial Extent and Richness," *IEEE Intern. Conf. Acoustics, Speech, Sig. Proc.*, vol. 3, pp. 2837–2840, May 2002.

[35] N. A. Goodman, "MIMO Channel Rank via the Aperture-Bandwidth Product," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2246–2254, June 2007.

[36] V. Raghavan and A. M. Sayeed, "Weak Convergence and Rate of Convergence of MIMO Capacity Random Variable," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3799–3809, Aug. 2006.

[37] V. Raghavan, R. W. Heath, Jr., and A. M. Sayeed, "Systematic Codebook Designs for Quantized Beamforming in Correlated MIMO Channels," *IEEE Journ. Sel. Areas in Commun.*, vol. 25, no. 7, pp. 1298–1310, Sept. 2007.

[38] V. Raghavan, V. V. Veeravalli, and A. M. Sayeed, "Quantized Multimode Precoding in Spatially Correlated MIMO Channels," *Submitted to IEEE Trans. Sig. Proc.*, 2008, Available: [Online]. `http://arxiv.org/abs/0801.3526`.

[39] V. Raghavan, A. M. Sayeed, and V. V. Veeravalli, "Low-Complexity Structured Precoding for Spatially Correlated MIMO Channels," *Submitted to IEEE Trans. Inform. Theory*, 2008, Available: [Online]. `http://arxiv.org/abs/0805.4425`.

[40] S. G. Srinivasan and M. K. Varanasi, "Constellation Design for the Noncoherent MIMO Rayleigh Fading Channel at General SNR," *IEEE Trans. Inform. Theory*, vol. 53, no. 4, pp. 1572–1584, Apr. 2007.

[41] S. G. Srinivasan and M. K. Varanasi, "Optimal Constellations for the Low SNR Noncoherent MIMO Fading Channel," *Submitted to IEEE Trans. Inform. Theory*, 2007.

[42] J. H. Kotecha, Z. Hong, and A. M. Sayeed, "Coding and Diversity Gain Tradeoff in Space-Time Codes for Correlated MIMO Channels," *IEEE Global Telecommun. Conf.*, vol. 2, pp. 646–650, Dec. 2003.

[43] C. Lin, V. Raghavan, and V. V. Veeravalli, "To Code or Not to Code Across Time: Space-Time Coding with Feedback," *To appear, IEEE Journ. Sel. Areas in Commun.*, Oct. 2008.

[44] V. L. Girko, *Theory of Random Determinants*, Kluwer, MA, 1990.

[45] E. Visotsky and U. Madhow, "Space-Time Transmit Precoding with Imperfect Feedback," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2632–2639, Sept. 2001.

[46] Y. Liang and V. V. Veeravalli, "Correlated MIMO Rayleigh Fading Channels: Capacity and Optimal Signaling," *Proc. IEEE Asilomar Conf. Signals, Systems and Computers*, vol. 1, pp. 1166–1170, Nov. 2003.

[47] V. Raghavan, V. V. Veeravalli, and R. W. Heath, Jr., "Reduced Rank Signaling in Spatially Correlated MIMO Channels," *IEEE Intern. Symp. Inform. Theory*, 2007.

[48] L. H. Ozarow, S. Shamai (Shitz), and A. Wyner, "Information Theoretic Considerations for Cellular Mobile Radio," *IEEE Trans. Veh. Tech.*, vol. 43, no. 2, pp. 359–378, May 1994.

[49] A. J. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity Limits of MIMO Channels," *IEEE Journ. Sel. Areas in Commun.*, vol. 21, no. 5, pp. 684–702, June 2003.

[50] V. Raghavan and A. M. Sayeed, "Multi-Antenna Capacity of Sparse Multipath Channels," *Submitted to IEEE Trans. Inform. Theory*, 2008, Available: [Online]. `http://www.ifp.uiuc.edu/~vasanth`.

[51] S. Verdú, "Spectral Efficiency in the Wideband Regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[52] I. S. Reed, "On a Moment Theorem for Complex Gaussian Processes," *IRE Trans. Inform. Theory*, vol. 8, pp. 194–195, Apr. 1962.

[53] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley, NY, 1st edition, 1960.

[54] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-Antenna Channel Hardening and its Implications for Rate Feedback and Scheduling," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1893–1909, Sept. 2004.

[55] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its applications*, Academic Press, NY, 1979.

[56] I. H. Woodhouse, "The Ratio of the Arithmetic to the Geometric Mean: A Cross-Entropy Interpretation," *IEEE Trans. Geoscience and Remote Sensing*, vol. 39, no. 1, pp. 188–189, Jan. 2001.

[57] G. Jasso, "Measuring Inequality: Using the Geometric Mean/Arithmetic Mean Ratio," *Sociological Methods & Research*, vol. 10, no. 3, pp. 303–326, Feb. 1982.

[58] H. Lanzinger and U. Stadtmueller, "Refined Baum-Katz Laws for Weighted Sums of I.I.D. Random Variables," *Stat. and Prob. Lett.*, vol. 69, pp. 357–368, 2004.

[59] R. A. Durrett, *Probability: Theory and Examples*, Duxbury Press, 2nd edition, 1995.

[60] P. S. Bullen, D. S. Mitrinovic, and P. M. Vasic, *Means and Their Inequalities*, D. Reidel Publishing Company, 1988.

[61] V. L. Girko, *Theory of Linear Algebraic Equations with Random Coefficients*, Allerton Press, NY, 1996.

[62] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, Oxford, UK, 3rd edition, 2001.